PROSODY AND SPEECH PERCEPTION

by

Jerzy Zdzisław Józef Kirakowski

A Thesis presented for the
Degree of Doctor of Philosophy
University of Edinburgh

December, 1977.

ABSTRACT

The major concern of this thesis is with models of speech perception. Following Gibson's (1966) work on visual perception, it seeks to establish whether there are sources of information in the speech signal which can be responded to directly and which specify the units of information of speech. The treatment of intonation follows that of Halliday (1967) and rhythm that of Abercrombie (1967). By "prosody" is taken to mean both the intonational and the rhythmic aspects of speech.

Experiments one to four show the interdependence of prosody and grammar in the perception of speech, although they leave open the question of which sort of information is responded to first. Experiments five and six, employing a short-term memory paradigm and Morton's (1970) "suffix effect" explanation, demonstrate that prosody could well be responded to before grammar. Since the previous experiments suggested a close connection between the two, these results suggest that information about grammatical structures may well be given directly by prosody. In the final two experiments the amount of prosodic information in fluent speech that can be perceived independently of grammar and meaning is investigated. Although tone-group division seems to be given clearly enough by acoustic cues, there are problems of interpretation with the data on syllable stress assignments.

In the concluding chapter, a three-stage model of speech perception is proposed, following Bever (1970), but incorporating prosodic analysis as an integral part of the processing. The obtained experimental results are integrated within this model.

D E C L A R A T I O N

The work reported in this thesis is my own.

CONTENTS

Abstract	ii
Declaration	iv
List of Tables	vii
List of Figures	ix
Acknowledgements	x
Chapter One	1
Three Theories of Speech Perception	2
Grammar	12
Prosody	17
Chapter Two	25
Experiment One	26
Experiment Two	44
Experiment Three	60
Experiment Four	72
General Discussion	78
Chapter Three	84
General Introduction	85
Experiment Five	102
Experiment Six	118
General Discussion	126
Chapter Four	134
General Introduction	136
Experiment Seven	145
Experiment Eight	151
General Discussion	165
Chapter Five	168
Introduction	169

The Model 179
Evidence 185
Appendix I: Experimental Materials 205
Appendix II: Additional Calculations 213
Bibliography 218

LIST OF TABLES

Table

2.1 Disambiguation Scores for the Two Speakers 36

2.2 Cross-Categorization of Ss' Responses to Each Sentence 38

2.3 Success at Disambiguation 38

2.4 Number of Sentences whose Readings Fell into the Six Possible Combinations 39

2.5 Average Height and Standard Deviation of Midpoints of Central Vowels 52

2.6 Means Summary Table for Experiment Two 54

2.7 Analysis of Variance on the Control Scores 55

2.8 Analysis of Variance on the Uncorrected Experimental Scores 55

2.9 Analysis of Variance on the Subtracted Correction Scores 55

2.10 Analysis of Co-Variance 56

2.11 Simple Main Effects Table from Subtracted Correction Data 56

2.12 Mean Words Correctly Reported in Experiment Three 63

2.13 Analysis of Variance Summary Table 63

2.14 Simple Main Effects Summary Table 65

2.15 Number of Words Reported Correctly at the Time-Compression Ratios 75

2.16 Regressions of Performance on Time-Compression Ratio 76

3.1 Total Probabilities of Correct Response, Experiment Five 107

3.2 Results of the Nine Analyses of Variance 108

3.3	A-Priori Tests for Experiment Five	109
3.4	A-Posteriori Tests for Experiment Five	109
3.5	Total Probabilities of Correct Response, Experiment Five, "One-Either-Way" Method	111
3.6	Analyses of Variance on Experiment Five Scored the "One-Either-Way" Method	114
3.7	A-Posteriori Tests for Experiment Five, Scored "One-Either-Way" Method	115
3.8	Average Probabilities of Correct Response, Experiment Six	125
4.1	Number of Inter-Word Intervals in which One Tap Occurred, Cross-Classified	147
4.2	Intercorrelation Matrix for Number of Syllables Reported per Segment	154
4.3	Totals of Disagreement Scores	158
4.4	Differences between Observed and Expected Disagreements	160
4.5	Intercorrelation Matrix for Stresses	167

LIST OF FIGURES

Figure

2.1 X' Scores (Subtracted Correction) for Experiment Two 57

2.2 Mean Comprehension Scores from Experiment Three 64

2.3 Mean Comprehension Scores from Experiment Four 74

3.1 A Flow-Diagram of Information in the Logogen Model 87

3.2 Average Probability of a Correct Response, Speech and Non-Speech Suffixes 90

3.3 Diagrammatic Summary of Morton's Explanation of the Stimulus Suffix Effect 94

3.4 Average Probability of a Correct Response, Experiment Five, "Serial" Scoring 106

3.5 Average Probability of a Correct Response, Experiment Five, "One-Either-Way" Scoring 113

3.6 Average Probability of a Correct Response, Experiment Six, "Serial" Scoring 122

3.7 Average Probability of a Correct Response, Experiment Six, "One-Either-Way" Scoring 123

ACKNOWLEDGEMENTS

I would like to thank Dr. T.F. Myers and Dr. J. Brown, my supervisors, for their technical and moral support; and my colleagues in the Speech Communication Laboratory--both present and past--for the many hours of discussions we have enjoyed together. I also thank the Edinburgh University Computer Allocation Board, and especially Mr. J. Ross, for their lavishness with computer time. Without the help of these good people I cannot see how this thesis would ever have reached the form in which it is now.

I am also grateful to the Medical Research Council for supporting my first two years' research, with whose assistance experiments one, two, three, seven and eight were carried out.

On a more personal note, I thank my wife, Susan, for her support, encouragement and assistance at sticky moments; and Tony Pinfold, Esq., for the services a good friend can only supply.

This thesis is dedicated to my dear brother, Józef Kirakowski. I have endeavoured to be worthy of the confidence he placed in me. Requiescat.

This letter, written by a friend of ours,
Contains his death, yet bids them save his life. (Reads)
Edwardum occidere nolite timere bonum est
Fear not to kill the king 'tis good he die.
But read it thus, and that's another sense:
Edwardum occidere nolite timere bonum est
Kill not the king 'tis good to fear the worst.

C. Marlowe
The troublesome raigne and
lamentable death of Edward the
second, King of England.
Act V, scene iv.

Chapter One

Three theories of speech perception

A rudimentary theory of speech perception might sound something like the following. An utterance is composed of a sequence of connected sounds. The perceiver hears these sounds, identifies them, and organizes them into words; groups of words into meanings. In order to hear the sounds, the perceiver must be able to segment them from the input; in order to identify them, the perceiver must be able to distinguish them from all other sounds. This theory has been shown to be wrong on many grounds, some of the most important of which will be summarised in the following paragraphs.

Most important of all are two reasons from an examination of the sounds of speech themselves why the rudimentary model will not work. The theory postulates that sounds are segmented and identified uniquely. Chomsky and Miller, (1963), discussed two conditions which must be met before we can assume this. One is the linearity condition: this states that sounds of speech must proceed in order; the end of one sound must occur before the start of the next. The other is the invariance condition: a particular sound must be analysable as one and only one class of sound. If neither of these two conditions are met, Miller and Chomsky claimed, then a theory similar to our rudimentary theory is insufficient.

One may take issue with the examples usually given of the violation of both the linearity and the invariance conditions, however, while agreeing with

Miller and Chomsky for other reasons. Firstly to deal with the linearity condition, although the phenomenon of co-articulation (c.f. Öhman, 1967) is often cited as a telling argument against the rudimentary theory the objection is more apparent than real. For instance, Liberman (1970) made the point that speech is not a "cipher" but a "code"; he meant that the device which perceives speech at the levels we are discussing resembles not so much a passive bank of filters, each attuned to a different phoneme, but rather an active device such as a grammatical parser, which extracts the complexly-embedded phonetic information in the speech signal much like a parser designed to deal with words and sentences may be able to extract deep structure sentoids which are complexly embedded in surface structure. Indeed, to assert that the linearity condition is not met simply because there is not a one-to-one relationship between the order of information in the speech wave and that in the percept is to ignore Gibson's dictum and warning to perceptual psychologists (Gibson, 1966, p. 242): "the fallacy that simple regularity is the only kind."

Furthermore, to claim that since all perceptually similar sounds are not produced at exactly the same bands of energy or multiples of it at exactly the same times, the invariance condition is not met, is to make a very special claim about the speech perceptual system that is not true, say, of the system of colour perception in vision or the perception of key in music, and doubtless any other perceptual system as well.

For instance, with colour vision, Land and

McCann (1970) have demonstrated that the perception of colour depends not on the "absolute" value of the luminance-vs-wavelength distribution but on our ability to register zones of discontinuity between adjacent patches of different colour ("relative"). Similarily, in musical perception, listeners are rare who can detect a modulation by considering the absolute pitch of the new keynote; a modulation is perceived by considering the relative distance between the new keynote and the old and the tonal path the composer took from the one to the other. Again, to quote Gibson (op. cit. p. 93):

> the interesting fact about the sounds of speech is that they are to a striking degree invariant with changes of pitch, loudness, and duration. The effective stimuli are relational, not absolute; ratios, not quantities...

Some of the more crucial reasons for agreeing with Chomsky and Miller are these. Firstly, with regard to the linearity condition, speakers, even such practiced speakers as announcers on the radio, sometimes miss out whole syllables from their speech without being noticed by well-practiced listeners listening to their native tongue (J. Brown, personal communication). Secondly, with regard to the invariance condition, particularly in fast speech, speakers can make all their vowel sounds so nearly alike that in a phonetic transcription, they can sometimes be described by one symbol. These perceptual observations have been studied experimentally: Warren (1970) showed that taking out a phoneme from a recorded sentence and replacing it with a swish of white noise of equivalent duration has a paradoxical effect on perception: the listener typically hears <u>both</u>

the extirpated phoneme and the swish of white noise; the white noise may even be heard at another subjective location in the sentence. It can also be shown that replacing all the vowel sounds with one single vowel sound in a recording of a passage has little effect on the comprehensibility of the passage.

A third argument is also often made against the rudimentary theory, which is based on some estimations made by George Miller. Calculating from the amount of time it takes for the human organism to make one decision, he concluded that the speed of decision-making in the brain cannot hope to match the speed of the phoneme: indeed, it was more likely that the unit of decision-making was of the order of about two or three words (Miller, 1962). This argument is weak on two counts: firstly, since listening is a skilled activity, one would expect that the organism might have evolved specific strategies for dealing with the rate--after all, skilled pianists, for example, can play notes faster than the most fluent speaker can articulate phonemes. To say that the pianist in this instance is only attending to phrases (of music) or other note groupings is not tantamount to a denial that he can (and does) also play each note individually. The counter-argument to Miller's point is that we do not know how the perceiver in this instance can "chunk" his decision-making, to use another of Miller's own arguments (c.f. Miller, 1956).

Secondly, since speech processing is not necessarily a serial activity, speakers can very well take advantage of the fact that information is often coded in parallel (c.f. our discussion of co-articulation, above) and process it in parallel. On the other hand,

this argument is supported by evidence which shows that constraining the input sequence by making it grammatical has a facilitating effect on comprehension, even when familiarity with the materials is accounted for (c.f. Miller, 1962). Precisely how grammar affects speech is still a matter for conjecture: in particular, these and other experiments (c.f. Fodor, Bever, and Garret, 1974, for an extensive review) fail to distinguish between what, to borrow terms from memory mechanisms, we may call the "proactive" and the "retroactive" effects of grammatical facilitation. That is, does grammar help us to chunk what has already been heard (retroactive) or does it help us predict what is going to be heard (proactive)?

Before we start on where these sorts of considerations might be taking us, let us take stock of the present state of the rudimentary model. Something else other than the sequence of phonemes is giving us information about what the speech means, since we can do without strict linearity or invariance at the phonetic level and hear well enough; and there is also something important to perception about units approaching the size of grammatical constituents.

To account for constancy in perception of speech, even when the speech was degraded to the extent of being inadequately invariant or linear, the speech "analysis by synthesis" model was proposed by Halle and Stevens (1962). Their expressed aim was to propose a device "which did not rely crucially on segmentation". By segmentation in this context, they meant, of course, segmentation into phoneme-like units.

Analysis-by-synthesis works on the principle

of a feed-back loop. Speech is perceived by being compared to an internally generated speech; a comparator sends mismatch messages to the generator in order to improve the quality of internal generation. What is "heard" is not the spoken stimulus, but the internal generation. The device incorporates an initial pre-analyser that looks at the incoming speech evidence in order to guide the guess of the generator. The activities of the generator are constrained in that it generates according to grammatical rules, and relies on context and expectation. Analysis-by-synthesis was taken seriously by psycholinguists of no mean standing (c.f. Chomsky and Halle, 1968, p. 24) and its popularity in some quarters seems unabated today (c.f. Lackner and Tuller, 1976).

To re-phrase Halle and Stevens, the motivation for the analysis-by-synthesis model was to present a device which would give a segmented percept of a characteristically un-segmented input. In order to do this, the model states that what is heard does not come from the speech input but from the activity of the internal generation. Thus in some discussions of the model, the word "hear" is enclosed in quotation marks, since what is "heard" is not the stimulus but the generation (see, for instance, Lieberman, 1967, p. 165).

There are two problems associated with this state of affairs. One is: how can the perceptual device keep track of the input--that is, how can the generator know it is keeping pace with the information it is meant to be matching (see also Fodor, Bever and Garret, 1974, pp. 317 - 319). For if a mismatch occurs,

then either of two things is possible: the generator has guessed wrong, or the generator is out of step. It can be shown that if the generator thinks it has guessed wrong, whereas it is really out of step, if the speed of message transmission in the model is finite, the attempts by the generator to produce the correct answer will ensure that the generator gets more and more out of step. If on the other hand, the generator thinks it is out of step, whereas really it has guessed wrong, no amount of searching of input before the moment it has come to this erroneous decision, or waiting for the input to catch up on the internal generation will free the generator from its quandary.

In itself, this is not a particularily damaging criticism of the model, since all that is required, is to say that some parts of speech obviously do help the generator keep up with input, and can be relied upon to do so. For instance, Neisser (1967, chap. 7) attributed precisely this function to stress and rhythm in speech. However, from this we are forced to admit that there might indeed be some features of speech which are produced regularly, and which we can identify without having recourse to analysis-by-synthesis. If this is the case, one may legitimately ask, firstly, how is this information extracted independentley of the generator-comparator loop, and secondly, is it possible that other aspects of speech may not also be handled in this fashion, thus largely obviating the need for the entire model?

The second objection stems from introspective evidence. We always hear the peculiarities of speech of people, who for instance are unfortunate enough to have a speech defect or an accent dissimilar to ours. The model, of course, would not predict this without

some alteration. Since what we hear is the internally-generated signal, any mismatch between our internal generation and the sound of the input would be put down to error: this in fact is not the case. A possible reply to such a criticism is that the generator can become attuned to specific distortions regularily imposed by the speaker on his speech: this however would imply that the comparator has some way of coding the qualitative nature of the mismatch. If we allow the comparator such a powerful facility, once again, the need for the generator to actually produce matches becomes less and less pressing. In a similar vein it might be asked of the model; how does it handle the times that we do produce slips of the tongue, misarticulations, etc. not of the systematic variety, but at random? Even granting the mismatch component the ability to handle defects of speech and accents the ability to perceive spontaneous speech errors presents a serious challenge to the model as it now stands.

In summary then, two arguments can be presented against the analysis-by-synthesis model both of which stem from the fact that the model is a "constructionist" type of model: it regards speech as impoverished data that has to be treated as stimulation rather than information; has to be added to rather than perceived directly. The question may legitimately be asked: are there sources of information in speech that can be responded to directly? An analogy with visual perception may be pertinent here. For many years, Gibson claims, it was considered that our perception of three dimensions had to be constructed out of the two two-dimensional representations of the world

at the retinae. Gibson's theory of three-dimensional perception asserts that there is no need to construct this representation: it is right there, in the ambient optic array (Gibson, op. cit.).

Another sort of explanation of the phenomenon of invariant perception in speech, despite the problems discussed above, concentrates on the meaning of the speech input. It is sometimes combined with a variety of the analysis-by-synthesis model which works to produce not actual phonetic sequences, but grammatical constituents, which it then matches in some unstated way with the speech input (see for instance, Wanner, 1968, Chap. 4). However, the identification with analysis-by-synthesis is arbitrary. The approach claims that speech need not be accurately specified in phonetic detail because very often, we know what is going to be said from the context and circumstances of the utterance.

In support of such an approach, Lieberman showed that the word "nine" is less carefully articulated in a sentence such as "a stich in time saves nine" where it is fairly easy to guess what the last word would be than in a sentence in which the last word could be any number at all, for example: "the next word you will hear is nine" (Lieberman, 1964).

This perception by deducing possible meanings no doubt does characterise one of the ways in which we can perceive speech: however, we do not talk in stereotyped phrases all the time--language is almost infinitely creative in potential at least--and the theory is silent as to how the listener does actually perceive the word "nine" in the second of the above

examples. Taken to extremes, of course, the whole attempt can be made to look risible: speech is either totally redundant, in which case there is no obvious point to actually talking unless one is addicted to the sound of one's own voice, or else is very non-redundant, in which case a listener has to attend to the sounds of speech. How he may be able to accomplish this is not given by the theory; that is, the processing rules which might have led to the level of perceptual representation assumed by this approach are not demonstrated (c.f. Myers, 1973, p. 3).

In summary so far, therefore, it seems that although the rudimentary model was not adequate for three very good reasons, two more sophisticated models, those of analysis-by-synthesis and perception by hypothesis-testing are also found wanting. What has motivated the incorporation of grammar into these models is firstly a realization that there may not be sufficient evidence in the phonetic aspects of speech to enable correct perceptual identification, and secondly, the knowledge that perception does seem to proceed in chunks roughly corresponding to small grammatical constituents. This latter is a finding which is not restricted to Miller's paper (Miller, op. cit.)--see Fodor, Bever and Garret (op. cit.). It is now necessary to examine some of the ways in which grammar may be incorporated into a model of perception.

Discussions of grammar cut across these three models of speech perception so far outlined. Grammar may be used with a "retroactive" effect in the rudimentary model; it is used predictively, thus proactively, in the analysis-by-synthesis model, and the hypothesis-testing model. In the analysis-by-synthesis model, it generates new matches for input spectra; in the hypothesis-testing model, it generates from the hypotheses the listener has in mind about what the speaker is going to say, some guess as to the words the speaker is going to use of has used to convey his intentions. Grammar may be incorporated into these models in three ways.

Firstly, a set of linguistic rules may be incorporated wholesale into a model. A well-examined example of this was what might be called the "derivational" theory of speech perception. This was based rather naively on linguistic descriptions of relationships between the so-called "surface" and "deep" forms of language. It identified the linguistic deep form with meaning; the surface form with input; and the transformations described by linguists were supposed to be directly isomorphic to the processes whereby the listener related surface to deep forms.

Summarising over twenty years of research which attempted to illustrate the perceptual reality of linguistic transformational rules, Fodor, Bever and Garret wrote in 1974:

> experiments which undertake to demonstrate

> the psychological reality of structural descriptions characteristically have better luck than those which undertake to demonstrate the psychological reality of the operations involved in grammatical derivations (p. 241).

In other words, there is scanty evidence to support notions of the psychological reality of grammatical transformations in performance models.

It is perhaps appropriate to restate the distinction made by Noam Chomsky, the doyen of transformational grammatical theory, with regard to competence and performance. The approach he advocated was aimed at elucidating the competence of the speaker/hearer in his native language: it is not concerned with how the speaker/hearer actually speaks or hears (see Chomsky, 1965, p. 9). Linguistic descriptions enable us to speak about language in a methodical way: it would indeed be a happy accident if a particular linguistic theory of transformations was also endowed with psychological reality. With the derivational theory of perception, this accident did not happen (see also Derwing, 1973, pp. 308 - 312).

Secondly, grammar may be incorporated rather more loosely into a series of rules whereby the listener may be said to analyse the incoming speech. An example of such an approach is the "augmented transition network" theory of speech perception (see Kaplan, 1972) which in effect postulates the existence of an in principle infinitely expandable and recursive network through which the perceiver can move as he analyses input,

checking predictions against input evidence, and back-tracking in cases of mismatch between prediction and evidence. An acknowledged defect of such a theory is the potential power of the basic concept of an augmented transition network which can recognise sentences as "grammatical" (i.e. they will receive an assignment from the network) which are yet perceptually unacceptable--for instance, sentences with multiple centre-embeddings.

Augmented transition networks are best left aside for the moment as frameworks within which a more specific theory of perception can be accommodated and implemented, for instance, on a computer. The third approach that will be mentioned can itself also be realised an as augmented transition network (see Kaplan, op. cit.).

This approach may be likened to some sort of template-matching with templates flexible, but of limited size. Bever (1970) and Clark and Clark (1977) discuss what we may call "perceptual mapping strategies". These strategies describe how best a listener may be able to relate deep to surface forms in a psychologically plausible way. These perceptual mapping strategies will be described in greater detail in the concluding chapter: they are interesting in that they do attempt to formulate the sort of information in the speech signal (considered here as a sequence of lexical items rather than sounds) that may specify meanings.

As an example of such a strategy we may take Bever (op. cit.):

> Strategy D: Any Noun-Verb-Noun sequence within a potential internal unit in the surface structure corresponds to "actor-action-object" (p. 298).

That is, any sequence which does not fit the requirements of the Noun-Verb-Noun template, does not correspond to "actor-action-object". Presumably, the model will try other templates if one fails.

The incorporation of grammar into this sort of model is pretty abstract: perceptual mapping strategies such as these have little or no relation to any current notions of linguistic transformational rules. The discovery of which are the relevant strategies and how are they ordered is left to the perceptual psychologist rather than being decided for him at the outset (c.f. the derivational theory). A set of such strategies may be likened to Watt's notion of an "abstract performative grammar" (Watt, 1970).

Although Bever (op. cit.) talks of "segmentation strategies" and Clark and Clark (op. cit.) distinguish between "syntactic" and "semantic" strategies, there seems to be no agreement, firstly, as to what the sources of information relevant to each of these processes may be, and secondly, the order in which they are supposed to take place. Thirdly, it is not clear how the perceiver can get from sound to the level of representation that these strategies demand without

having to do some analysis which relies upon there being sufficient information in the sounds of speech. The first two questions will be discussed in the concluding chapter; in the following section, an attempt will be made to propose a solution to the third.

Prosody

It will be taken as axiomatic that grammar and meaning are specified (to use Gibsonian terminology again) by the sounds of speech: the sounds of speech are specific to grammar and grammar is specific to meaning, where "specific to" is considered equivalent to "conveys information about".

To assert that speech is perceived in terms of sequences of phonemes, is, to use another visual analogy, like saying that perception consists of the integration of patches of variously-coloured light. This is wrong: in vision, we clearly perceive not patches of light but forms and shapes. There is nothing magical about our perception of these forms and shapes--we see them because they are a property of visual input. In the same way, one may argue, the perception of speech does not consist of integrating sequences of phonemes together: we perceive meaningful regularities directly. Once again, there should be nothing that we have to add or do to the sound wave in order to produce sentences and phrases, there should be adequate information in the signal to specify them.

The suggestion we shall adopt is that the phenomena of intonation and rhythm correspond to the invariants in the sound of speech which enable a listener to perceive the meaning of the utterance directly. Together, intonation and rhythm may be said to define the prosodic contour of the utterance.

In this final section, therefore, the linguistic and phonetic bases of prosody will be discussed, and some necessary terminology will be introduced. Although the relationship between prosody and grammar will be touched upon, psychological and psycho-linguistic aspects of prosody will not be reviewed here because the subject matter at present lacks a unifying thread; instead, such studies of prosody will be examined in the introductions to the relevant experimental chapters.

With regard to intonation, the treatment will follow that of Halliday (1967), so some introduction to his system is necessary. Speech, according to Halliday, may be considered as a sequence of "tone groups", roughly corresponding in extent to the grammatical clause. There are three important aspects of the tone group that will concern us. Firstly, a tone group may be considered "inclusive" and subject to the linearity condition. That is, once the boundaries of the tone groups have been decided, any smaller segments of speech (e.g. words or syllables) are either in one or another tone group, never in more than one at a time or partly in one and partly in another. In fact, it seems that the average span of a tone group is seven to eight syllables (c.f. Laver, 1970), and it is supposed to have a number of boundary markers associated with it (see chapter four for a more complete discussion of boundary markers).

Secondly, any tone group possesses a characteristic shape or intonational contour,

and thirdly, the contour draws attention to one (or in some cases two, but perhaps not more than two) particular syllables. These syllables are called "tonic" syllables, and if they form part of a poly-syllabic word, the word is called the "tonic item" of the tone group. Tonic syllables can be cued by a certain sort of inflection of fundamental frequency (see the discussion of stress, below) and Halliday classifies five sorts of inflections that can characterise the tonics of English tone-groups. In the normal course of events, the tonic falls on the last content word of the tone group (again, see the discussion of stress): this constitutes what is called the "unmarked" case. In special circumstances, a "marked" tonic will fall on another lexical item. The tone-group can thus be divided into pre-tonic, tonic, and post-tonic parts. Post-tonics do not, according to Halliday, have any contrastive function in English; the voice simply carries on in the direction given by the intonation at the end of the tonic. There are several pre-tonic contours associated with each tonic (see ibid, pp. 16 - 17).

The part of the sound wave that gives rise to the perception of the intonation contour is identified with the fundamental frequency of vibration of the larynx (see Lehiste, 1970, chap. 3). Halliday does not, unfortunately, give any hard-and-fast phonetic criteria for determining the extent of the tone-group. From the discussion of the marked vs. unmarked distinction with regard to the number and place of the tonic, it follows, for example, that it is not simply a matter of waiting for the tonic syllable and then placing a boundary marker before the next lexical item.

Before considering the rhythmic contour of the tone-group, the notion of stress, which links intonation and rhythm, must be examined. Stress as will be used here is a property of a word: all words are capable of receiving stress, although it turns out that some words rarely do. Thus one division of words is according to the probability of them receiving stress: a typical distinction along these lines is that between "content" and "function" words. Content words include nouns, verbs, adjectives and adverbs; function words include articles, prepositions, conjunctions and auxiliary verbs. Function words rarely receive stress, content words invariably do; function words are mostly monosyllabic, but no such constraint is found with regard to content words. In content words, stress does not seem to fall on prefix or suffix parts of the word.

The acoustic manifestations of stress have been throrougly dealt with by Lehiste (1970, Chap. 4). They seem to be, in order of decreasing importance: shift in fundamental frequency, amplitude, duration, and voice quality.

Since duration is one of the cues to stress, combining stressed and unstressed syllables will mean that the onset time of each syllable will not appear regularly (at least in English). The combination of long and short syllables will produce a rhythmic pattern which relates closely, through the fundamental frequency cue to stress, and through the connection between tonic syllable and stress, to

the intonation contour. We will follow Abercrombie (1967) and recognise that there are two levels of stress: ictus, which is a stressed syllable, and remiss, which is unstressed. Ictus and remiss syllables are combined into what Abercrombie calls "metrical feet"; the composition of a foot is one ictus, followed by any number (including zero) remiss syllables. Thus, by definition, a foot must and can only contain one ictus. Tone groups are however occasionally encountered which begin with a couple of unstressed syllables: such groups of syllables should be regarded as feet with a silent ictus (Abercrombie, op. cit.).

One interesting feature of feet in English is that they tend to be perceived as occurring at roughly equal temporal intervals (unlike syllables in English). It is not known whether this is due to a regularity in the speech signal itself, or whether this is due to the registration of salient syllables in a particular way in memory: difficulties with locating the perceptual centre of a syllable (see, for instance, Morton et. al. 1976) make a study of this phenomenon difficult. Synthesised speech which has been produced so as to make the stressed syllables occur at objectively exactly equal temporal intervals sounds very strange and unnatural indeed on account of this invariant metronome-like beat. Furthermore, there seems to be a trading relationship between the length of a foot and the number of unstressed syllables inside it, and the position of the foot within the rest of the tone group (see Lehiste, 1973).

The relationship between the tone group and grammatical features of speech has also been noted by linguists other than Halliday. For instance, Trager and Smith (1951) talked about the "phonemic clause": this for their system was what determined the size of the grammatical clause, and seems to be equivalent to our notion of the tone group. Chomsky and Halle (1968) described a set of rules which related in a very specific way the surface structure of a sentence and the "stress contour" associated with it. For purposes of discussion, we may equate the stress contour with the selection of tones in Halliday's system (see Bolinger, 1972a, pp. 49 - 51). Chomsky and Halle's rules enable one to deduce which lexical item, for instance, will receive the "primary stress" of the sentence. It is clear that this is deducible not only from the surface structure, as they claim, but that considerations of meaning will also play a part in the assigning of major stress.

Halliday (op. cit.) simply mentions that the tonic gives the "informational focus" of the utterance without further specifying what this informational focus may relate to.

The Trager and Smith system was fairly severely criticised by Lieberman (1965) who demonstrated by means of an experiment that the phonemic pitch levels and the terminal junctures of the Trager and Smith system often had no distinct physical cue in the speech signal. The Chomsky and Halle system, one may say, has made itself fairly immune to this kind of treatment by claiming that very often, the levels of stress

predicted by the system have no physical counterparts in the signal, and are supplied by the listener by means of analysis-by-synthesis. The theory was nevertheless given an incisive review by Bolinger (1972b) whose title neatly summarises the point of his criticism: "Accent is predictable (if you're a mind-reader)".

These claims and counter-claims are of more concern to the linguist than the perceptual psychologist: they have been brought in to demonstrate that it seems to be a well-held contention between linguists of fairly different persuasions that intonation has some sort of connection with grammar, although opinions as to the specificity of such a connection and how it is best represented in a linguistic system, of course, vary.

While the intonation of a tone-group may be considered, to draw on a musical analogy, as the spoken analogue of the succession of tones of a melody, the rhythm of a tone group may be likened to the rhythm of the tune (see for instance Longuet-Higgins, 1976). The units of a melody which contain the pitches and carry the rhythms by virtue of their length and degree of accentuation, are of course notes. It could be considered that the equivalent units for speech are syllables. This is not to claim that the syllable is <u>the</u> unit of speech perception but that many prosodic phenomena are best described in relation to the syllable, just as a lot of musical phenomena may be described in relation to the note.

Consideration of the way the simile breaks down is instructive: in Western music, at least, the notes of the tune remain fairly stable with regard to pitch while the note is being sounded; whereas the intonation contour on each syllable of spoken speech characteristically is moving. In the same way, the metrical scheme of a melody which is all written to the same time-signature will be fairly regular in execution by a competent performer (or at least, departures from regularity will be for effect rather than on principle) whereas it is not at all clear that the metrical scheme of speech has regularity to the same extent. Finally, in music, notes (unless connected by a device such as a portamento or glissando) are fairly distinct from each other. This may not be the case in speech in that it may be difficult for two listeners to agree where (for instance in terms of constituent phonemes) one syllable begins and the other ends.

From this analogy, it should follow that a distinction should be made between the tone group when considered as a succession of tones or contours, irrespective of rhythm, and another for the tone group when considered as intonation and rhythm: the words "intonation contour" and "prosodic contour" seem perfectly appropriate. The question to which the research to be reported in the following chapters addresses itself to is, to what extent can the prosodic contour of an utterance be considered as the invariant specific to meaning?

Chapter Two

The four experiments to be reported in this chapter are designed to investigate whether prosody has any real effect on the perception of speech. The chapter after this will produce some evidence relating to the locus (in a functional rather than an anatomical sense) of prosodic processing of speech; and the last experimental chapter will demonstrate the sorts of prosodic regularities that do exist in spontaneous speech.

The first experiment to be reported in this chapter will deal with the problem of whether, in the case of some ambiguities that cannot be resolved by recourse to lexical parsing strategies, prosody has any effect on the perception of meaning. The next three examine the effect prosody has on the perception of material not usually considered ambiguous: experiment two, on a passage from an essay; and experiments three and four on short sentences.

Experiment 1

An ambiguous sentence is one that may have more than one plausible meaning. Although the majority of sentences we use may be ambiguous in one way or another, the fact is that these ambiguities are hardly ever noticed, and the alternative interpretations are often highly implausible. On the other hand, there are some sentences which we can readily accept as ambiguous.

A useful scheme for classifying ambiguous sentences was proposed by MacKay and Bever (1967).

Their system of classification has the advantages that it fits in well with a widely-held linguistic theory (Chomsky's 1965 Extended Standard Theory) and that it has been used frequently since in psycho-linguistic research on ambiguity. Sentences are classified according to where in the process of generating a structural description of the sentence the question of ambiguity arises: in the meanings of the individual words ("lexical"); in the surface structural description ("surface structural") or in the deep structural representation ("deep structural").

It must be said that this scheme has been severely criticised by Garcia (1976) whose major contention seems to be that in Chomsky's theory, all ambiguity is represented at the deep structural level. No doubt MacKay and Bever would wish to argue that it does not matter: surface structural ambiguities would be represented both at deep and surface structure; lexical ambiguity would be represented both at deep structure, and the lexical insertion levels (see the reply by Bever, Garret and Hurtig, 1976). A further difficulty with the classification is that allocating a sentence to one of these categories is by no means a cut-and-dried procedure (see, for instance, Mehler, and Carey, 1967, footnote 3).

Nevertheless, this classification does satisfy an intuition about grammar that has existed since the time of Aristotle at least (see for instance his "De Sophisticis Elenchis"). What is particularly

interesting for the purposes at hand is the category of ambiguous sentences whose ambiguity is ascribed to a plurality of possible surface structural descriptions. If prosody is important in perception; if intonation should be regarded as part of the grammar of English (see Halliday, 1967) and if prosodic cues to phrase structure exist and function just as cues from function words, closed class morphemes and word position do (see Neisser, 1967, pp. 259 - 267), we would expect that their presence would most clearly be felt in the case of ambiguity at surface structure--or "amphiboly".

Rather than use the somewhat more clumsy locution "ambiguity at the surface structural level" we may employ an old-fashioned term: "amphiboly". An amphibolous sentence is one that can have two and no more grammatical constructions. Amphibolous sentences are therefore ideal for forced-choice discrimination tasks: the bread-and-butter of perceptual psychology. "Ambiguity" and "amphiboly" will be used as deemed appropriate by context.

A number of attempts have been made to demonstrate the influence of prosody on the perception of meaning in ambiguous sentences. Unfortunately, the picture is still confused. Bolinger and Gerstman (1956) and later Lieberman (1967) found that appropriately placed portions of blank tape could produce disambiguations of the recorded phrase "a light house keeper", regardless of the intonation of the spoken sentence (it can either mean "a man who

tends a light-house" or "a house-keeper who does not weigh much". The interpretation of these experiments is open to some dispute--see chap. four).

Nash (1970) performed an experiment using synthesised intonation contours for the phrase "John likes Mary more than Bill" (which can mean either that "John likes Mary more than Bill does" or "John likes Mary more than he likes Bill"). The fundamental frequency contour was manipulated and two groups of Ss were asked to make judgements as to whether Bill was the subject or object of the sentence represented by any given production. There is equivocal evidence in Nash's data for an intonational disambiguation effect.

Scholes (1971) asked Ss to judge the location of the "sentence break" and "most stressed word" in the phrase "good flies quickly", recorded as "the good flies quickly passed" and "the good flies quickly past" by five trained readers. The relative amplitude of the words "good" and "flies" was found to be the major determinant of "sentence break" (sc. location of the major syntactic boundary) and stress judgements. This finding is in direct contrast to the work reported by Lieberman (op. cit.).

Lehiste (1973) presented four informants with fifteen ambiguous sentences, taken promiscuously from linguistic discussions of ambiguity. The informants read each sentence once, the ambiguity was pointed out to them, and they then read it again

twice, trying to bring out the two different meanings. Thirty Ss then attempted to identify the meanings intended by the speakers. Lehiste found that the disambiguations which depended on a difference of surface structure bracketing were fairly successfully cued by pauses introduced by the speakers at the relevant parts of the sentence. She does not give data separately for the first reading as compared to the second and third.

These results can at least in part be explained by Lieberman's theory of speech as an "optimal code" (Lieberman, op. cit.). He argued that a speaker will not use intonational contrasts to cue a disambiguation unless there are no other means available to him, for speech is a code which contains as little information as is necessary for the communication of intended meaning. If an ambiguity can be settled by context, there is no need, so Lieberman's argument runs, for the speaker to signal it by intonation (if this is indeed possible). Thus sentences which are spoken with the intention of producing a contrast will show the desired disambiguation-by-intonation effects. However, in the normal course of events this is not necessary, nor will speakers do it unless specifically asked.

Toner (1975) reported an extensive experiment on the power of intonation to resolve ambiguity. There are two aspects of his study that are particularly relevant to the present discussion. One is that he attempted to discover whether intonation

had any effect on the resolution of deep, surface, and lexical ambiguity with the categorization as described above. His results show that intonation had most effect on disambiguating amphiboly, but it was a small effect (four sentences out of ten in this condition showed a "significant effect of disambiguation by intonation", and no sentence in any of the other two conditions did at all). The other interesting feature of Toner's experiment is that he attempted to test Lieberman's "optimal code" hypothesis. If speakers recorded an ambiguous sentence embedded in a context paragraph, they should, under Lieberman's hypothesis, be less likely to signal any dismabiguation of the sentence by intonational means than if they recorded the sentence in isolation, without a context paragraph, with the different meanings of the sentence pointed out to them.

Four informants recorded the stimulus sentences, firstly embedded in a suitably disambiguating context and then without any context, contrastively. Contexts were edited out, and the resulting readings presented to Ss in a forced-choice discrimination task.

Now, given that the effect of intonation overall was insignificant, Lieberman's theory would predict a significant interaction between ambiguity type and condition of recording. The interaction is predicted because one type of ambiguity (surface structural) should be disambiguated better by recordings from one condition (contrastive recording). Since we

know that the only sentences that achieved a disambiguation-by-intonation effect were of the surface structural ambiguity type, any significance in this interaction should therefore be attributable to the manner in which the sentences of this condition were recorded. However, this interaction turned out not to be significant, either. Since less than half the sentences in this condition showed any significant effect of disambiguation-by-intonation anyway, perhaps this insignificant interaction is not really surprising.

From Toner's description, it appears that his informants might have been reading contrastively in the non-contrastive condition; furthermore, although an attempt might have been made to control for contrast effects in recording, contrast effects in listening were not controlled. Thus listeners heard both versions of each ambiguous sentence and thus might have been responding to the second reading partly on the basis of their response to the first. From what one may gather from Lehiste's paper (op. cit.), these same criticisms apply to her.

The present experiment attempts to establish whether in principle it is possible to disambiguate amphiboly by speaking, and secondly, if it is possible to do this in principle even when the informant and listener is naive. The phrase in principle is used since the phenomenon would hardly be less real if demonstrated with one informant than with four. If it is established, it is then possible to extend the experiment using a larger sample of informants (none of the reported research on this topic to date has ever

used more than four--in one case, five--informants).

Method

Design. Nineteen amphibolous sentences were obtained using as sources MacKay (1966), MacKay and Bever (1967), and Bever, Garret and Hurtig (1973) as cited by Garcia (1976). All but two were of the surface structural type. On two, our analysis disagreed with MacKay's in that there was an obvious difference in the surface structural bracketing which reflected the two meanings the sentences could have (sentences two and seventeen).

An example of an amphibolous sentence is:

1. Beethoven, knowing how great symphonies sound, composed nine.

This has two meanings:

2. Beethoven knew how good a symphony sounds, so...

and

3. Beethoven knew what all great symphonies should sound like, so...

Two contexts were then generated for each sentence, each cueing a different meaning. For example, the two contexts generated for (1) above were:

4. We perhaps imagine that a composer with his eye on what the public wants is a twentieth-century phenomenon, probably on the pop scene. Beethoven, knowing how great symphonies sound, composed nine. How well they sold was proof of how well they were liked.

and

5. Beethoven, knowing how great symphonies sound, composed nine. It didn't seem to worry him that Mozart had written nearly five times as many that were also great. Hadyn had written over a hundred symphonies.

When possible, this cueing was also reinforced by supplying appropriate punctuation marks.

The resulting 38 paragraphs were then randomised, with the constraint that no amphibolous pairs succeeded each other. An informant then recorded all the paragraphs. At the end, she was asked whether she had noticed anything special about the paragraphs. She replied that "some of the words kept on cropping up" but registered surprise when the embedded ambiguous sentences were pointed out to her. It may be concluded that the informant was truly naive as to the purpose of the experiment.

The 19 sentences were also recorded by E, who tried to bring out the different meanings as clearly as he could by the use of prosodic contrasts. These were the non-naive recordings.

The naive recordings were then edited so that only the amphibolous sentences were preserved and the recorded contexts were discarded. The two sets of recordings were then randomised yet again to make up four sequences, two from each speaker. The recording of each sentence was presented twice in succession (see procedure). No sentence was repeated within a sequence, to avoid the possibility of Ss picking up meanings on the basis of contrasts.

Ss were then instructed to listen to one of the four sequences and to make forced-choice discriminations as to which meaning had been intended by the speaker for each sentence (thus each S heard only one recording of each sentence and heard only one speaker's recording).

Procedure. The paragraphs were printed on small cards and handed one by one to the naive informant. E recorded the (contrastive) stimuli in pairs. The recordings were made on a Revox two-channel tape recorder, using a moving-coil monocardioid microphone. Stimuli were played back to the Ss on a PYE 9137 model recorder. Ss were tested three or four to a group; twenty Ss listened to each sequence. Each S had a printed form in front of him which explained briefly the two possible meanings for each sentence. S was instructed to listen to each stimulus, read both appropriate paragraphs of explanation on the form in front of him, listen to the stimulus (again), and then to make his response by indicating whichever meaning he thought the speaker intended for the recording.

Subjects. The naive informant was a female postgraduate student in the psychology department. She had an English accent similar to E's. The eighty Ss were undergraduates at Edinburgh University, participating in the speech communication practical class.

Results

For each recording, the number of correct responses, scored according to the following criterion, was obtained. A response was deemed correct if the S indicated the meaning intended by the speaker. The raw data and the sentences used are reproduced in the appendix. The total number of correct responses for each amphibolous sentence was obtained (i.e. summed over both readings for the two speakers separately). The maximum score possible was 40: this indicated complete agreement by all the Ss on the two meanings intended by the speaker for that sentence. If the responding was on a chance basis, the score would be 20. The obtained means and standard population distributions are summarised in table 2.1. A "t" test was used to determine whether the obtained means came from the same distribution as predicted by a null hypothesis of just guessing the responses. The null hypothesis was rejected at the 1% level of confidence for both speakers.

TABLE 2.1: Disambiguation scores for the two speakers,

	naive speaker	non-naive speaker
Mean	26.42	31.52
s.d.	6.058	4.234
t	4.496	11.540

d.f. = 18; $p<.0005$ with a one-tailed test. Mean expected by chance = 20.

It will be noticed that the average correct responses for the non-naive speaker's recordings is higher than that for the naive speaker's, although

both scores are significantly above chance level. The difference between the means is significant at the 1% level of confidence, using a "t" test for uncorrelated samples on the data reported in table 2.1 (d.f. = 34; t = 2.927; p .01). This difference suggests that contrastive recording has a stronger disambiguating effect.

A clearer picture will emerge if the scores for each reading are compared one by one between the two voices. Each score may fall into one of three states:

SH: there was a significantly higher proportion of correct responses;

NS: there was no significant difference between the number of correct and incorrect responses; and

SL: there was a significantly higher proportion of incorrect responses. "SH" (Significantly higher) implies that the reading facilitated the correct response. "NS" (Not significant) and "SL" (significantly lower) imply either that the speaker did not cue the meaning adequately, or that the meaning was incapable of being cued by reading. SL implies the existence of a bias in favour of the alternative meaning, whether it be due to the speaker's reading or the listener's perception is impossible to tell. Using chi-square, it was found that correct scores higher than 15/20 could not be expected by chance at the 5% level of confidence.

Comparing readings between the two speakers, we find that nine combinations of state possibilities

exist into which any pair of readings may fall. Table 2.2 shows the obtained frequencies for each combination of states.

TABLE 2.2: Cross-categorization of Ss responses to each sentence, between the two speakers, in terms of categories SH, NS, and SL.

non-naive speaker:		SH	NS	SL
naive speaker:	SH	19	0	0
	NS	8	8	0
	SL	1	2	0

See text for an explanation of the categories SH, NS, and SL.

As can be seen from the above table, it was never the case that the naive speaker managed to communicate a meaning more successfully than the non-naive speaker. Table 2.3 shows this clearly.

TABLE 2.3: Number of times when the non-naive speaker was more, just as, or less successful at producing a disambiguation response than the naive speaker.

event	frequency
non-naive better	27
both same	11
naive better	0

Since the expected frequencies are higher than five in table 2.3, "chi-square" can be applied to examine the differences between these proportions. The differences are significant at the 5% level (d.f. = 2; chi square = 9.125).

Finally, the non-naive recording contained more sentence pairs which were both successfully cued, and it was never the case that the naive informant managed to cue a pair of meanings that the non-naive informant failed to. Table 2.4 shows this data.

TABLE 2.4: Number of sentences whose readings fell into the six possible combinations of states SH, NS, or SL for each speaker.

combination	naive speaker	non-naive speaker
SH & SH	5	10
SH & NS	7	8
SH & SL	2	0
NS & NS	4	1
NS & SL	1	0
SL & SL	0	0
TOTALS	19	19

It is meaningless to analyse this table statistically without collapsing cells, and all the meaningful collapses have been demonstrated.

Conclusions

Three points emerge clearly from the foregoing analysis: (1) prosody obviously does help the disambiguation of amphiboly; (2) the disambiguation effect is stronger when the speaker is consciously attempting a contrast prosodically, but it still exists when a speaker records other (contextual) cues

that a listener might use; (3) there seems to be a bias towards one meaning in many amphibolous sentences. The case of both readings of a sentence producing a non-significant result is rare (N = 5/38, see table 2.4).

Although conclusion (2) might be confounded with the effect of having two speakers (the experimenter might have been better at signalling disambiguations by prosody, anyway), the important point is surely that the disambiguation scores for the naive speaker are significantly above chance.

Discussion

Although the experimental hypothesis has, on the whole, been verified, there are only fifteen (out of a total of 38) pairs of readings that gained an "SH & SH" rating (i.e. both meanings of the sentence successfully cued). Of these fifteen, five were from the naive informant's readings, and ten were from the non-naive informant's. Another fifteen pairs of readings gained an "SH & NS" rating (i.e. only one meaning of the sentence successfully cued--see table 2.4).

Whereas the "SH & SH" pairs can be considered as strong evidence for the support of the hypothesis, the "SH & NS" sentences require some explanation. Two hypotheses offer themselves. Firstly, listeners could, in principle, pick up the correct cues, but it so happened that neither of the two informants

used in the study were particularily good at producing such information. As evidence which could possibly support such an explanation, of the sentences which gave the naive informant's sixteen readings classified as "NS", eight were given "SH" and eight "NS" classifications in the non-naive informant's readings: that is, the non-naive informant was able to improve on half of them, but it was never the case that the naive informant was able to improve on any of the non-naive informant's readings (see table 2.2).

Secondly, it could well be that in some of these cases, prosody was in principle incapable of cueing the desired meaning: that is, just as lexical ambiguities do not seem to be resolvable by prosody, neither do some surface structural ambiguities. Most probably, both explanations are true and apply to different sentences: this does, however, make the interpretation of results from studies which employ ambiguous sentences rather difficult.

This difficulty dogs the assessment of the strength of lexical parsing strategies independently of prosody within the present design. It may be argued that there are strong strategies of this kind which interact with prosody--for instance, a "SH & NS" pair could, by lexical parsing strategies alone, be a "SH & SL" pair (same meaning given by most listeners to both readings). The results of prosodic processing would agree with the lexical parsing strategy for the meaning of the "SH" reading, but contradict it in the case of the other member of the pair. As a result,

this reading would be categorised as neither "SH" nor "SL" but as something in between--"NS". Alternatively, lexical parsing strategies could be relatively powerless, and all that the "NS" score signifies is that for this meaning, prosody was not very powerful as a source of information either. In fact, there are only two such "SH & SL" pairs, both in the naive informant's readings. One becomes a "SH & NS" and the other a "SH & SH" pair in the non-naive informant's readings.

An appeal to measurements taken from Ss reading the stimuli of the experiment and writing down the most plausible meaning they can think of, or that which comes first into their heads, as estimates of the parsing strategy effect uninfluenced by prosody (c.f. Toner's work, op. cit.) is fallacious on two counts. Firstly, the requirements of English grammar and orthography demand that for some of the ambiguous sentences, when written down on paper, cues be given from sources such as punctuation and spelling which would immediately prejudice the issue as to which meaning was being perceived.

Secondly, and more generally, such an appeal suumes that silent reading is equivalent to listening to prosodyless speech. This assumption may well be invalid: for instance, readers may employ a strategy specifically for reading, which says:

> in the absence of any other information from spelling, punctuation, or context, assume that the meaning intended by a written sentence is that which would be intended

by the sentence if it were spoken on an unmarked intonation contour.

On account of these considerations, discussion of the importance of prosody for speech perception as well as the status of Lieberman's "optimal coding" hypothesis will be deferred till the next three experiments have been reported.

Experiment 2

The simplest demonstration of the importance of prosody for the perception of speech would show that speech which was in some way restricted in its range of prosodic features would be more difficult to perceive than "normal" speech. In fact, this experiment has been carried out four times to our knowledge, and the combined overall results are, at best, ambiguous.

Glasgow (1952) tested the comprehension of poetry and prose delivered in "good intonation" and "monopitch" styles of delivery. The comprehension test was a multiple-choice questionnaire. He found that comprehension was adversely affected by the monopitch delivery, both for poetry and for prose. Of the two groups of Ss involved in his study, however, one group's results are totally insignificant while the other group's are highly significant, which suggests that only one group is responsible for this result. One may wonder why the other group did not also provide significant results; Glasgow does not mention any of this. A re-analysis of Glasgow's data is contained in the appendix.

Diehl, White and Satz repeated a similar experiment in 1961; but their conclusion is in direct contradiction of Glasgow's:

> the use of interval and inflection does not affect listener comprehension (p. 67).

Thomas (1969) performed an experiment in which "longish" sentences were read to listeners. Upon hearing a sentence, Ss were presented with a list of words. Within this list was one word that had actually occurred in the middle of the preceeding sentence, as well as a number of similar-sounding words which had not. Ss' task was to indicate the word that had occurred. Sentences were read in "normal" intonation, "monopitch" intonation, and "monotone" intonation: this latter description refers to intonation characterised by

> falling pitch patterns and qualities necessarily judged subjectively as boredom, fatigue, disinterest, lifelessness and dullness (p. 111).

Thomas concluded that there was no significant effect of mono-pitch intonation, but monotonous intonation did adversely affect Ss in the recognition task.

Stowe and Hampton (1961) made some interesting observations in the course of an experiment which investigated the intelligibility of sentences constructed out of pre-recorded words and syllables (their research interest at the time was the synthesis of speech). The segments from which the stimuli were constructed were pre-recorded at two speeds--fast and slow. They concluded:

> embedding a word in a constructed sentence, where the pitch, stress and phrasing are not properly rendered, decreases the probability of recognition...if more processing time is allowed the listener, he can nevertheless overcome (this) lack (p.811).

In other words, negative findings in the research cited above could be explained on the grounds that

Ss found the experimental tasks too easy. These considerations suggest that an important control is one for ease of attending; later we will have a chance to examine this question in the light of some data (in experiment four).

By contrast, a well-established influence on the comprehension of speech is the degree of grammaticality of the speech stimulus (c.f. Miller, 1962) which could be taken as evidence for the existence of lexical parsing strategies. It would therefore be an interesting question to compare any decrease in comprehension due to the manipulation of prosody with the already established decrease associated with the absence of grammatical structure.

Directly comparing the level of comprehension of prosodyless speech with ungrammatical speech is beset with conceptual problems. It is difficult to consider speech entirely devoid of prosody. In fact, following from the definitional problems discussed in chapter one, one may well ask, what is prosodyless speech? It would be reasonable to conclude that any comparison along these lines will depend simply on the severity of the prosodic reduction. A study which fails to demonstrate similar average comprehension scores for prosodyless, grammatical speech and ungrammatical speech may well be superseded the next day by one which, working to a severer criterion of prosodic reduction, succeeds.

Another way of posing the problem is to ask whether reducing prosody in ungrammatical speech

has the same sort of perceptual effect as reducing prosody in grammatical speech. This comparison also involves a conceptual difficulty: can ungrammatical speech ever have "normal" prosody? Although such a case may occur spontaneously only in a pathological syndrome, linguists do habitually consider some elements of prosody, at least, independently of the words of speech. If one can meaningfully consider these units of prosody abstractly, sundered from the words, then another can surely consider them artificially wedded to different, albeit less suitable words.

There are two possible predictions about the outcome. One follows from Bever, Fodor and Weksel (1965):

> though pauses, stresses, sound units, intonations etc. (sc. prosody) are heard as objective features of the flow of speech, it appears that these percepts are in fact, the result of some elaborate manipulation of the acoustic data.

They proceed to assign a crucial rôle in this manipulation to grammatical structure, and go on to say:

> beyond doubt, extra stresses or artificially induced pauses can induce structure in a random sequence.

That is, they would predict that since we perceive prosody by "elaborate manipulation of the acoustic data" there should be little difference between the grammatical conditions in which prosody is varied. On the other hand, they might predict that since "adding prosody" to an ungrammatical

stimulus induces structure within it, the effect should be to increase the comprehensibility of such a stimulus. An alternative prediction, on the other hand, would go something as follows.

The interaction between prosody and grammar is profound (see chapter one). Therefore, grammatical stimuli without prosody should be less intelligible than grammatical stimuli with prosody. On the other hand, strings of unrelated words ungrammatically arrayed have no such relationship with prosody. Therefore we would expect little or no perceptual effect in manipulating prosody in this case.

Both predictions would result in a significant statistical "interaction" term between prosody and grammar. Only analysis of the simple main effects would show which hypothesis was verified.

There is, of course, a third prediction (no statistical interaction term) and many more variants of the two hypotheses outlined above. These particular hypotheses have been selected because it was possible to find a theoretical motivation for both: no doubt a *post hoc* model can be found which will enable us to account for any combination of data.

In order to minimise memory load, a shadowing task was used: that is, *S* had to repeat what he had just heard out aloud while listening to what was being said to him after that. This paradigm raises several vexing problems; Chistovich *et al.* (1960) discuss perhaps the most vexing, which is that

different Ss will have different strategies for coping with the demands of the task. To control for this, all the Ss will shadow one introductory passage first, the same for all experimental conditions. Subsequent measurements on each S could then use this measure as a baseline, with the assumption that Ss' strategies will not change drastically during the course of the task.

Method

Design. The design was factorial, grammar and nonsense by normal and reduced prosody, with an introductory control passage that was the same for all four conditions. A "grammatical" (and meaningful) passage and an equivalent "ungrammatical" (nonsense) passage were devised. Both were recorded twice: once with "normal" prosody, and once on a monotone. These were the four experimental recordings. Each recording was prefaced by another grammatical, meaningful passage read with normal prosody.

All the recordings were mixed with white noise, whose level was constant throughout the experiment, and which was established beforehand to produce a shadowing success rate (on average) of 70% for the introductory passage.

Ss were randomly allocated to one of the four conditions. Their task was to shadow the two passages (the introductory and the experimental). A record was kept of the number of words correctly shadowed.

Procedure. (1) Obtaining the materials. A passage of approximately 90 words was selected for the absence of proper names and complicated grammatical constructions. This was the grammatical passage. The order of the words was reversed to produce the ungrammatical passage. Thus two stimulus paragraphs were constructed, each with the same words. One stimulus paragraph had the words in a grammatical and meaningful order. A similar passage was selected for the introductory recording. Both passages are reproduced in the appendix.

(2) Recording the materials. The monotone recordings were made by recording groups of ten words at the rate of 90 words a minute, and then splicing the groups together to make an uninterrupted recording which would have otherwise been impossible on account of the need to pause for breath. E (who recorded the stimuli for this experiment) kept the pace constant by means of a metronome tick played back to him through a pair of headphones. The ticks did not get picked up by the recording microphone. The preparation of the normal prosody recordings was more elaborate. The grammatical passage was recorded, and the manner in which it was spoken was analysed for: (a) stressed syllables; (b) pauses or breaks between sentences or phrases; (c) the shape of the intonation contour within each phrase. The ungrammatical passage was then divided up into "phrases" on the following scheme: if the grammatical reading contained nine words in its first phrase, the first "phrase" of the ungrammatical passage

was also made up so as to contain nine words. This procedure was applied to the next phrase, and so on. The intonation contour was then also (approximately) transferred to the "phrase" from the readings taken from the grammatical reading. All the words which had been stressed in the grammatical reading were still stressed in the ungrammatical script. The relative position of the tonic word (counted in words from the beginning of the phrase) was also transferred. The resulting script was practised and eventually recorded.

The readings were recorded on a Revox two-channel tape recorder and played back through a PYE 9137 tape recorder with a lateral response unit providing a white noise mix to a headset worn by S.

To check the accuracy of the recordings, the pitch level in the middle of the vowel for the first 34 syllables was measured from a mingographic recording of the fundamental frequency of the four recordings and also the introductory, control recording. The means and standard deviations of these measurements are given in table 2.5. They are given in millimetres above the resting level of the pen (which was constant for all the mingograph recordings). The resting level corresponded to a frequency of about 60 Hz., and one millimetre corresponds to about eight Hz. near the line for the resting level. As can be seen from inspection, the monotone recordings were markedly narrower in the range of fundamental frequency employed.

TABLE 2.5: Average height and s.d. of midpoints of central vowels in the first 34 syllables of the five recordings employed in experiment two.

recording	mean	s.d.
introductory	11.24	4.76
normal prosody		
grammatical	9.00	4.28
ungrammatical	6.67	4.01
monotone		
grammatical	7.67	0.95
ungrammatical	6.73	1.50

Data given in millimetres above resting level of pen.

(3) Ss' responses. Two experimenters scored the responses, ticking off on a script each word that S got right. A faint signal of the speech stimulus only (without white noise) was delivered through another headset on the table in front of the experimenters, and this helped them keep pace with their S. A word was counted as correct if it occurred in the right grammatical form not more than three words after the word had reached S's ears. Scoring was done in "real time" (i.e. while S was shadowing) after it was found by checking some of the sessions with the aid of a second recorder which recorded what S said that Es were accurate in their transcription.

Subjects. Ss were 68 undergraduate students, participating in the speech communication practical class. They were naive as to the purpose of the experiment.

Results

There are two alternative procedures recommended for taking into account Ss' performance in the control condition when analysing experimental results (c.f. Winer, 1971, pp. 752 - 754). One is a simple analysis of variance performed on a subtracted correction, the other is a (more complex) analysis of co-variance. The present data was analysed chiefly by the subtracted correction method, although an analysis of co-variance main effects table was also computed and found to be equivalent in all respects to the analysis of variance main effects table.

Each S's score on the experimental variable was corrected with reference to his score on the control variable by the following formula:

$$X' = C - X + 33$$

where X' is the corrected score for the S, X his score on the experimental variable, and C his score on the control variable. A constant of 33 was added to all the scores in order to eliminate some negative numbers. Table 2.6 shows the means of the four conditions for (a) the control passage; (b) the uncorrected experimental passages and (c) for the subtracted correction as given by the formula, above. Part (d) also gives the corrected means from the analysis of co-variance method. Since most attention will be devoted to the subtracted correction scores (part c), these are displayed graphically in figure 2.1.

TABLE 2.6: Means summary table for experiment two.

Condition	Control scores (a)	Uncorrected scores (b)	Subtracted scores (c)	Covariance method (d)
normal prosody				
grammatical	34.76	38.18	29.59	38.5
ungrammatical	31.28	7.89	56.39	11.5
monotone				
grammatical	44.00	31.47	45.53	28.2
ungrammatical	38.00	17.76	53.24	17.4

A two-by-two factorial analysis of variance was carried out on the means of (a), (b) and (c); the same statistical model was incorporated into the analysis of co-variance model. Tables 2.7 to 2.10 show the results of the main effects.

TABLE 2.7: Analysis of variance on the control scores.

Source	d.f.	SS	MS	f
Grammar	1	660.9	660.9	1.825
Prosody	1	1009.5	1009.5	2.788
Grammar x Prosody	1	0.9	0.9	0.003
Residual	64	23173.1	362.1	
Total	67	24884.5	370.8	

TABLE 2.8: Analysis of variance on the uncorrected experimental scores.

Source	d.f.	SS	MS	f
Grammar	1	8162.1	8162.1	43.198*
Prosody	1	38.8	38.8	0.202
Grammar x Prosody	1	1144.7	1144.7	6.058*
Residual	64	12092.7	188.9	
Total	67	21437.8	320.0	

TABLE 2.9: Analysis of variance on the subtracted correction scores.

Source	d.f.	SS	MS	f
Grammar	1	5131.69	5131.69	26.41*
Prosody	1	704.73	704.73	3.63
Grammar x Prosody	1	1571.48	1571.48	8.09*
Residual	64	12435.75	194.31	
Total	67	19843.65	296.17	

* Significant at or beyond the 1% level of confidence.

TABLE 2.10: Analysis of co-variance.

Source	d.f.	SS	MS	f
Grammar	1	5905.6	5905.6	55.575 *
Prosody	1	80.2	80.2	0.755
Grammar x Prosody	1	1113.2	1113.2	10.476 *
Co-variate	1	5398.0	5398.0	50.798 *
Residual	63	6694.6	106.3	
Total	67	19191.7	286.4	

We may summarise the information presented in these tables as follows: there is a moderate amount of interaction between the co-variate and the experimental variables, although the means of the co-variate, taken on their own, show no significant differences. The points of similarity between the analysis of co-variance and the subtracted correction analysis of variance are striking. The simple main effects table is shown only for the subtracted correction method (table 2.11).

TABLE 2.11: Simple main effects table from subtracted correction data.

Source	d.f.	SS	MS	f
Prosody on:				
grammatical	1	2160.02	2160.02	11.4 *
un-grammatical	1	89.94	89.94	0.45
Grammar on:				
normal prosody	1	6279.75	6279.75	33.33 *
monotone	1	504.75	504.75	2.60
Residual	64	12435.75	194.31	

* Significant at or beyond the 1% level of confidence.

From this table we may deduce the interaction can be explained as follows: when the stimulus passage is meaningful, and grammatically organised, there is a strong reduction in Ss' performance with reduction of prosody. When the stimulus is meaningless and ungrammatical, prosody has little or no effect on performance. All the significant results are significant beyond the 1% level of confidence.

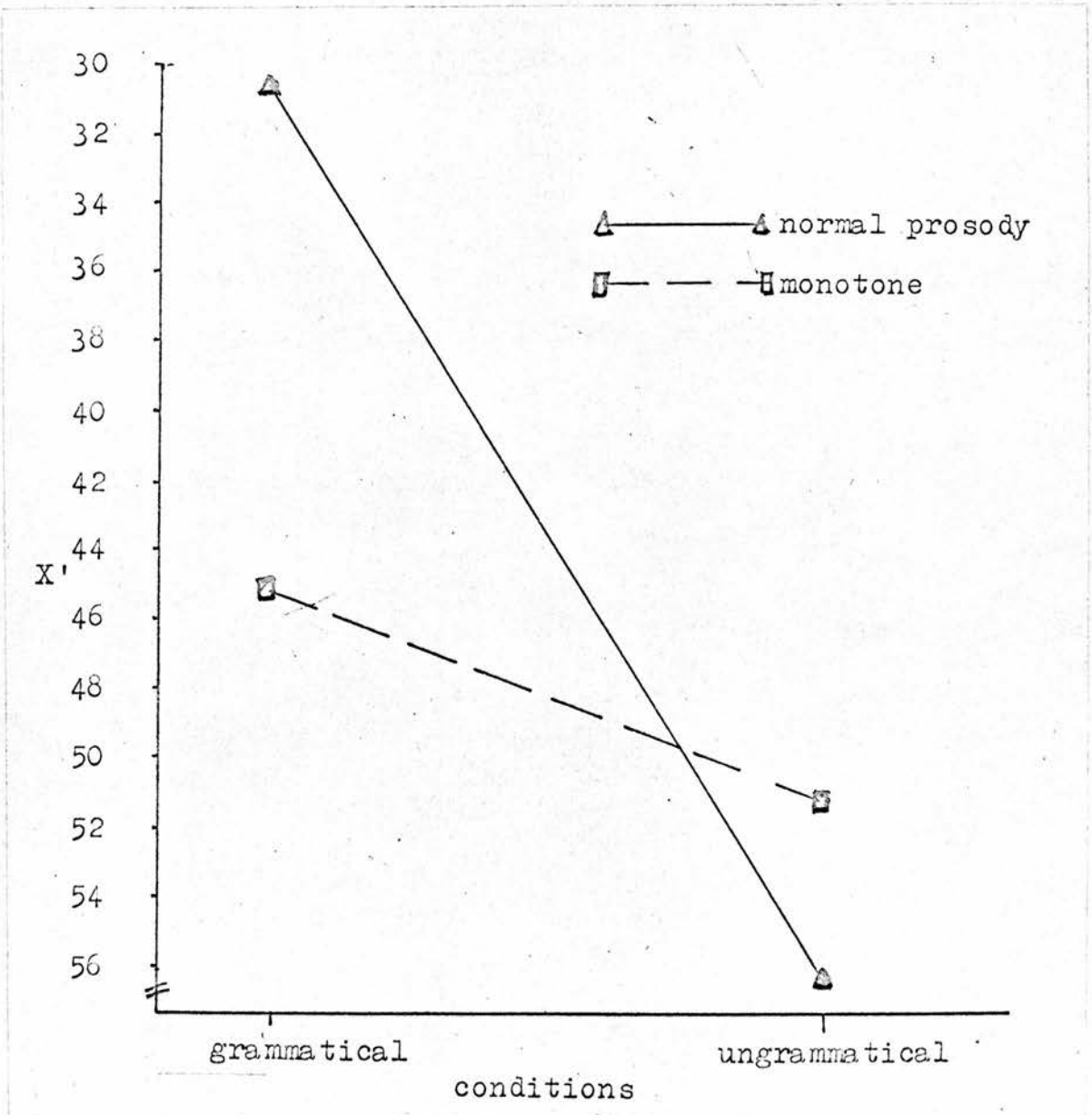

Fig. 2.1. X' scores (subtracted correction) for experiment two.

The preliminary trials run to establish the precise level of white noise interference needed to produce responding at the desired level in this experiment, and the amount of time compression in the next was done with postgraduate Ss from the psychology department, none of whom admitted to doing this sort of experiment before. When the time came to test the Ss who were all undergraduates it was found that the response rate was definitely lower on average (47% as against the postgraduates' 70%). The hypothesis that this was due to the equipment on which the preliminary trials were run being badly calibrated can be cautiously discounted as the effect appeared in two different experiments, using different sorts of apparatus. Other explanations are (1) lower motivation on the part of the undergraduate Ss, (2) the possibility that somehow the particular verbal skills needed for this experiment increase dramatically with age, or (3) the small sample of postgraduates used.

Vexing as this phenomenon is, the effects of it are not as far-reaching as they would have been had the data collected included the responses from the preliminary trials, or if a less homogenous sample of Ss had been used.

Discussion

The interaction model which is supported by the present evidence maintains that in the normal communication situation, the listener can use either his knowledge of the grammar or his knowledge of the

interrelationship between English prosody and grammar (or both) to help him follow what is being said. Crystal (1975, chapter 8) makes a case for prosody and grammar distributing the load of processing cues between themselves, but Stowe and Hampton's data (op. cit.) suggests that with a little effort, grammar can compensate for the lack of prosody. It is only when the difficulty of the task of listening is increased (as in the present experiment) that the perceiver may not have time or space enough to do the computations necessary to deduce the grammatical structure from the more monotone presentation.

Thus there are a number of respects regarding which the present experimental demonstration might be improved and extended. Firstly, the point would be clearer made if Ss had a task which gave them less choice of strategy: therefore control for variability between Ss could be taken case of in this respect at least in the design rather than post hoc in the analysis (see Winer, loc. cit.). Secondly, since the shadowing task is one that has been interpreted as demanding a considerable investment in memory space by the S (see Salter, 1973), a better demonstration might be carried out in a situation which makes fewer demands on memory. Thirdly, it would be interesting to examine the perceptual effects of different sorts of prosodic reduction.

One important thing the present experiment has shown us is the perceptual effect of prosody in connected speech. It remains to be seen what the effect is in shorter segments of speech.

Experiment 3

In the introductory chapter, the various components of prosody were discussed. Whereas in the two experiments already described prosody was dealt with as if it were an indivisible, unitary phenomenon, (c.f. Crystal, 1975, p.11), the time has now come to disentangle some of the components.

Specifically, this experiment seeks to determine whether: (1) the intonational contour has an effect separate to that of the rhythmic organization of speech; (2) the rhythmic periodicity of stressed syllables has any advantage over equal periodicity of all syllables ("stress-timed" and "syllable-timed"); (3) any periodicity of syllables has any advantage over no periodicity at all. It also continues the investigation, begun in the previous experiment, on the effect of prosodic features on ungrammatical nonsense strings.

Method

Design. The design was factorial: grammar and nonsense by four levels of prosody (see below). Each S received all of the grammatical and the nonsense strings, but different Ss were assigned to the different "levels" of prosody.

The grammatical stimuli were twenty five-word sentences. The ungrammatical stimuli were an equal number of five-word nonsense strings. The four levels of prosody were:

1. "Full prosody"--normal reading.
2. "Foot timed"--monotone, but keeping stressed-syllable isochrony.
3. "Syllable timed"--monotone, but retaining syllable isochrony (as in a "machine-gun" language--see Abercrombie, 1967).
4. "Spliced"--made up of pre-recorded words spliced together.

The order of the grammatical and ungrammatical strings was randomised within each level of prosody.

Procedure. Ungrammatical strings were made up by scrambling the order of the words in each grammatical sentence. The stimulus sentences and ungrammatical strings are listed in the appendix. The procedure for recording the ungrammatical strings for the first two conditions of prosody was essentially the same as in the previous experiment: that is, words which were stressed in the sentences were also treated as stressed in the ungrammatical strings. The intonation contours read in each sentence were also copied in the strings. In the foot timed and syllable timed conditions, isochrony of feet in one case and syllables in the other was maintained by speaking in time to a metronome click audible through headphones only. Spliced condition stimuli were made up from a tape of words pre-recorded in isolation, read on a monotone. The same recorded words were used for making up both grammatical and ungrammatical stimuli, so any difference between these two conditions will reflect the way the words are grouped rather than the way they are spoken.

In general, any sentences or strings which were unintelligible to a panel of four listeners at normal playing speed were recorded again. The stimuli were then subjected to time compression of 55% of normal playing time by means of a Lexicon VARISPEECH. This level was preset on the basis of responses from another panel of five listeners to the stimuli in the normal intonation ("full prosody") condition. This level of time compression on the VARISPEECH introduces a certain amount of transient noise perceived as a buzz superimposed on the recording. However, since the purpose of this procedure was to increase perceptual difficulty, and since this sort of interference was constant over all the conditions, it was not considered that it would detract from the purpose of the manipulation.

S's task was to listen to each sentence once, and then to write down what he had heard.

Responses were scored word by word. A word was counted as correct if it was in the correct grammatical form and in the right order in the sequence. In scoring the ungrammatical strings, homo-phones were treated as correct (e.g. "sea" and "see" were both correct responses to "see" in the string "to no-one came her see").

Subjects. All the recordings were made by E, the nine judges were postgraduates from the psychology department, and the 90 Ss were undergraduate students doing second year psychology.

Results

Table 2.12 shows the mean number of words reported correctly in each of the eight conditions. They are displayed graphically in figure 2.2. The appropriate statistical treatment for this data is two-factor experiment with repeated measures on one factor. Kirk's recommendations (Kirk, 1968) for calculation were followed. Tables 2.13 and 2.14 present the results of the analysis of variance and simple main effects, in that order.

TABLE 2.12: Mean words correctly reported in the eight conditions of experiment three.

	Grammatical	Ungrammatical
Full prosody	70.95	24.05
Foot timed	65.46	27.62
Syllable timed	56.67	25.52
Spliced	43.00	24.17

TABLE 2.13: Analysis of variance summary table.

Source	d.f.	SS	MS	f
A (prosody)	3	5406.46	1802.15	5.75*
Ss within groups	86	26911.75	312.93	
B (grammar)	1	50637.65	50637.65	1227.58*
A x B	3	4584.94	1528.31	37.05*
B x Ss within groups	86	3547.32	41.25	

* Significant at or beyond the 1% level of confidence.

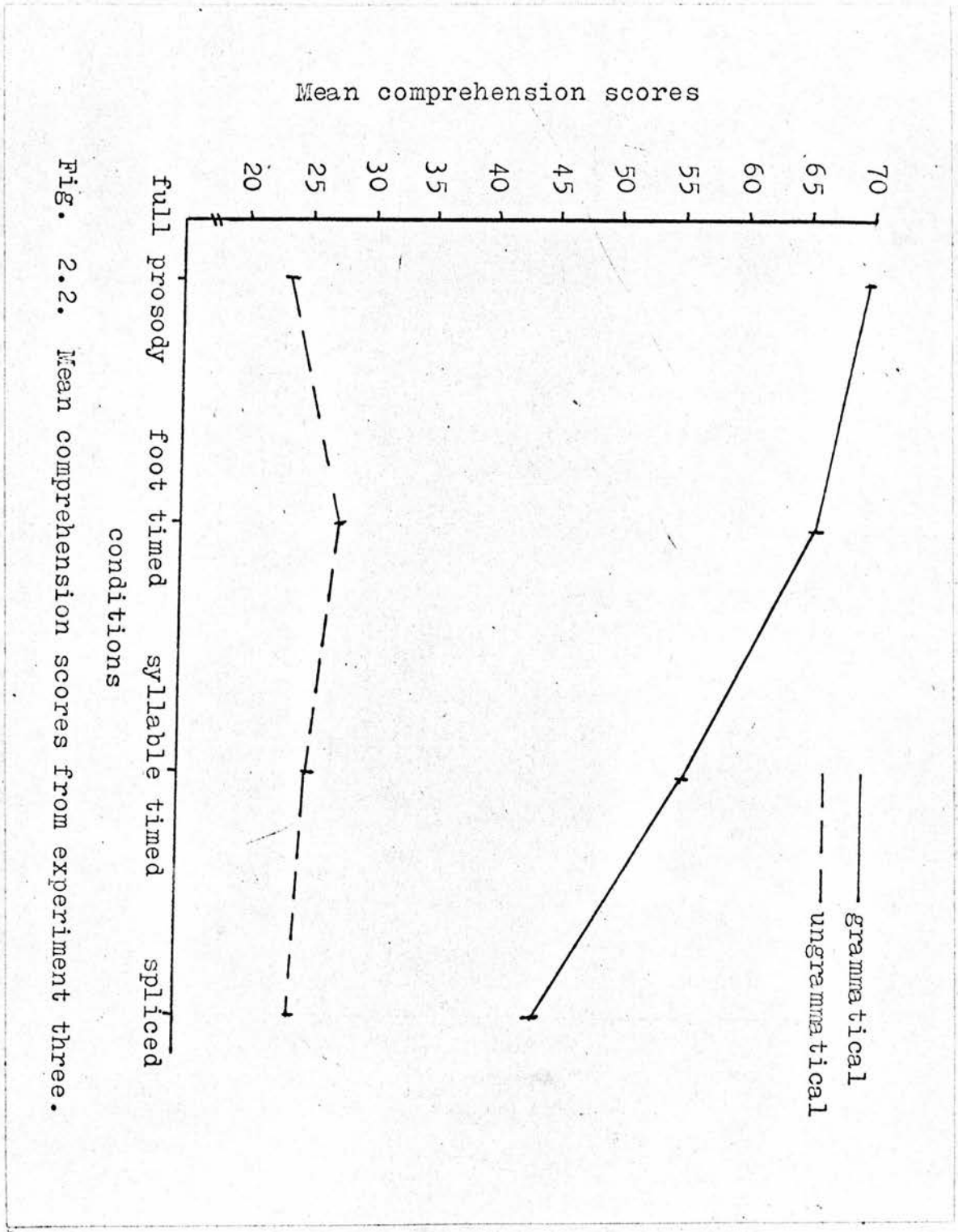

Fig. 2.2. Mean comprehension scores from experiment three.

TABLE 2.14: Simple main effects summary table.

Source	d.f.	SS	MS	f
Prosody on:				
grammatical	3	9958.51	3319.50	18.74*
ungrammatical	3	194.03	64.68	0.365
Grammar on:				
full prosody	1	23831.27	23831.27	577.72*
foot timed	1	17176.33	17176.33	416.39*
syllable timed	1	10183.71	10183.71	246.88*
spliced	1	4075.85	4075.85	98.81*
Prosody x Grammar	3	4584.94	1528.31	37.04*
Grammar x Ss within groups	86	3574.32	41.25	

* Significant at or beyond the 1% level of confidence.

This analysis confirms what can be seen from figure 2.2: that is, the effect of varying prosody on ungrammatical strings was insignificant, and the effect of varying prosody on the grammatical strings was significant. The effect of the grammatical-ungrammatical distinction was always significant. In addition, there was a strong interaction effect. All significance levels are beyond the 1% confidence limit. The fact that a five-word sequence should be well within most Ss' memory spans re-inforces the perceptual effects hypothesis discussed above.

Discussion

There are a number of substantial differences between experiments two and three; the number of prosodic levels in experiment three is not the least of them. However, one striking similarity is the shape of the graphs of the results, and the general statistical conclusions drawn from both experiments. This is a good argument for the reliability of the effect, especially since both experiments were carried out on large groups of Ss, using different sorts of measures of comprehension.

One methodological difference between the two is in the length of the stimuli used. Experiment two used a passage of about 90 words; this experiment used five word sentences and ungrammatical strings. Thus whatever else might have been the same between the experiments, experiment three had one sort of information invariably present--that of a silence at

the end of a sentence or string, in every condition. This of course was not true of the reduced prosody conditions of experiment two. As will be seen in the investigations to be reported later (experiment seven) such pauses might well be important cues to grammatical constructs. This phenomenon is the basis of the demonstrations reported by Bolinger and Gerstman (1957) and Lieberman (1967) discussed earlier. It may explain why the difference between means in the monotone conditions in experiment two is not significant, whereas the difference between the means of the spliced conditions in the present experiment is highly significant.

The difference between the full prosody and the foot timed conditions is an interesting one: it would seem from a consideration of these two conditions that intonational variation has a special function with regard to grammatical sequences. There are at least three sorts of information that intonation may be able to provide.

The first two follow from Halliday's treatment of intonation, and have already been mentioned in chapter one. The location of the tonic in Halliday's system determines, according to him, the intonational focus of the utterance (Halliday, 1967, p. 38); the choice of tone can be most usefully regarded as expounding the secondary systems of mood (referable to the primary terms of the mood system: declarative, interrogative, imperative; see ibid p. 40). Thirdly, some recent studies by Cutler (1976) suggest

that the intonation contour of the early part of the tone group may allow listeners to predict the shape of the contour of the later part. In a monotone rendition of a sentence, all this information is lacking; if the material does not make sense anyway, none of this "feed-forward" information is of any use.

To summarise these points: the important functions of intonation may be threefold. One is to signal the intonational focus, another to signal the mood, and the third to predict subsequent events. Needless to say, none of these three sources of information are particulairly relevant to the case of ungrammatical speech.

The difference between the "foot timed" and the "syllable timed" conditions may be regarded as primarily one of pacing the flow of information reaching the S. Lehiste (1974) and Bond (1971) found that in a click reporting experiment, reaction times to clicks located at stressed syllables are longer than reaction times to syllables located at unstressed syllable positions. If the reaction time to a click is proportional to the amount of time spent processing the syllable on which it happened to be superimposed, this finding implies that the amount of processing time each syllable needs can be to some extent discovered from a consideration of the stress pattern of the speech. From the theory of isochronicity of stressed syllables (e.g. Abercrombie, 1967), this could mean that the speaker times his utterance to ensure that arrival times of items which

important to the understanding of a sentence (i.e. content words, to refer once again to chapter one) can be predicted at least to a rough extent by the listener. Ladefoged and Broadbent (1960) made a point similar to this in their original report of the "click" experiment.

If even the isochronicity of stressed syllables disappears from the speech input, however, the listener's task will be that much more difficult. This question of predictability will arise again in a consideration of the "syllable timed" and "spliced" conditions (see below) but it is important to the point here that in meaningless material, the perceptual import of one syllable is most probably as great as that of any other. Certainly, the distinction between content and function words has no meaning in this case, since there is no meaningful content and no grammatical function.

Finally, some comment should be made about the "syllable timed " and "spliced" conditions. In experiment two, the monotone conditions did have a degree of regularity (in that it is possible to predict the arrival time of each word). In experiment three the "spliced" condition was composed of words of different durations, making the perceptual effect that of a markedly less regular utterance. It would be reasonable to attribute the decrement in performance between these two conditions to factors of timing and predictability. Unfortunately, it is not possible to compare effects between experiments in this case, other-

wise it would have been interesting to substantiate this last point by a comparison between the relevant conditions of experiments two and three.

Mills and Martin's theories about the perceptual coherence of items produced in the same utterance (see Mills and Martin, 1974) would predict that for ungrammatical strings, somewhere between the "full prosody" and the "spliced" conditions a significant decrement in performance should be noticeable. In fact, the effect of prosody on ungrammatical strings is not significant. Mills and Martin's theory explains data from experiments where strings of seven digits have been presented to Ss, and it is possible that in the present experiment the task was too easy for the beneficial effects of stimulus coherency to appear. This question will be discussed in greater length in the next chapter when the process of attending to speech information will be examined more closely.

One criticism that might be levelled at the interpretation of the results as they stand is that the lack of effect of prosody in the ungrammatical conditions is attributable to a "floor" effect--given that it is difficult to follow ungrammatical sequences of words anyway, the (constant) level of interference was so high, that in the ungrammatical conditions Ss would have been hard put to it to do any worse.

What we need to show, therefore, is that the pattern of performance obtained in the previous two

experiments holds good for less severe levels of interference. It was decided to take the stimuli of experiment three and to replicate some of the conditions at different levels of time compression ratio.

Experiment 4

The most extreme prosodic contrasts in the previous experiment are those between the "normal prosody" and the "spliced" conditions. They thus define the upper and lower limits of the observed phenomenon. If it is true that in general the lack or presence of prosody makes no difference to the perception of meaningless speech, then this pattern of equivalence should hold good over all ratios of time compression, and not just at 55%. On the other hand, as the task becomes harder due to a decrease of the time compression ratio (from 100%--normal playing time--to around 50%--half the normal playing time) we should expect that the beneficial effects of prosody should make themselves apparent in the perception of meaningful sentences.

Method

Design. The "spliced" and "full prosody" tapes of the previous experiment (which consisted of twenty grammatical sentences and an equal and equivalent number of ungrammatical strings) were played back to Ss at 100%, 90%, 80%, 70%, and 60% of normal playing time. This made ten conditions in all, and a fresh group of Ss listened to each tape in each condition (to avoid effects of familiarity with the presented material).

Procedure. Ss were asked to listen to each sentence in turn and to write it down immediately they had heard it. Three Ss were assigned to each condition, and the Ss of each group were tested together.

Subjects. Subjects were 30 second-year undergraduate psychology students, none of whom had been involved in any of the previous experiments, who volunteered their services.

Results.

The total number of words correctly reported by each S in each condition was determined according to the criteria laid down for the previous experiment, and the resulting means are displayed in table 2.15. Figure 2.3 displays these results graphically. It will be seen that the two grammatical conditions produce slopes which differ from each other, and also from the slopes of the ungrammatical conditions. These latter ungrammatical condition slopes are fairly similar. An examination of the regression coefficients shows this trend very clearly (the means of the relevant conditions from experiment three are included in these and all subsequent calculations on the data from this experiment). The regression coefficients are shown in table 2.16.

The experimental hypothesis can be demonstrated most clearly by the following method, which correlates differences between conditions with time compression ratios, using Pearson's r (see Runyon and Haber, 1968).

With regard to performance in the ungrammatical conditions, the hypothesis predicts no difference at any of the time compression ratios. Any difference actually observed between the means at any one ratio

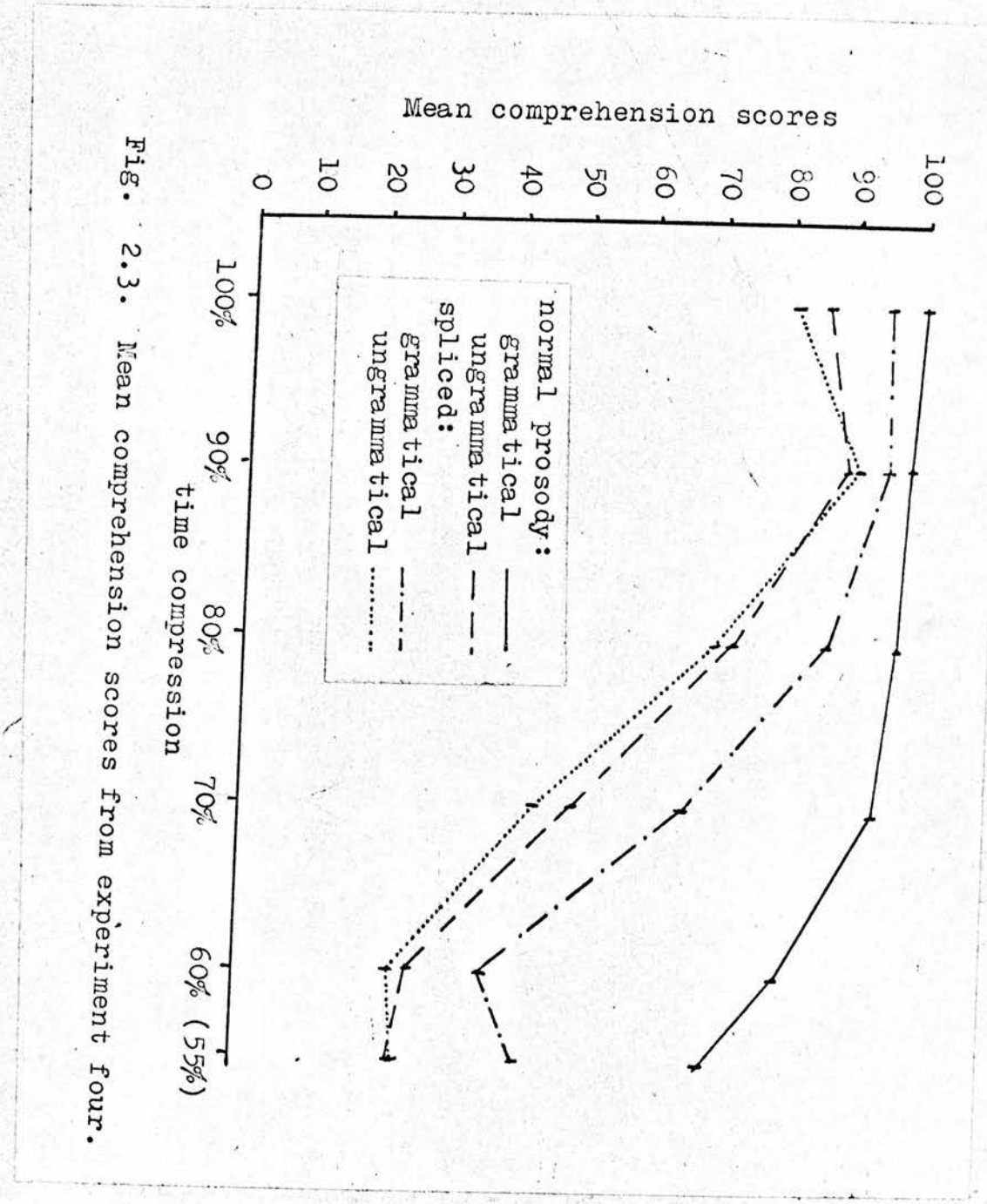

Fig. 2.3. Mean comprehension scores from experiment four.

TABLE 2.15: Average number of words reported correctly for each condition in the five ratios of time compression studied.

	100%	90%	80%	70%	60%	(55%)
Normal prosody						
grammatical	100.00	99.00	98.67	95.33	81.00	(70.55)
ungrammatical	86.00	89.00	73.00	50.67	26.67	(24.05)
Spliced						
grammatical	95.33	96.33	87.67	67.67	37.67	(43.00)
ungrammatical	81.33	90.00	71.33	44.67	23.00	(24.17)

Data from experiment three is shown in brackets for comparison.

is therefore attributable to chance. From this follows that the correlation between time compression ratios and differences between means of these two conditions should be near to zero. The actual calculated correlation is .06 and the regression is .01. The correlation coefficient is insignificant.

TABLE 2.16: Regressions of performance on time compression ratio.

Normal prosody	
grammatical	.6056
ungrammatical	1.5942
Spliced	
grammatical	1.4115
ungrammatical	1.5843

On the other hand, differences between performance in the grammatical conditions should increase as the task becomes harder and prosody extends its benefits. In this case we should expect a significant negative correlation between differences and ratios of time compression (as percentage playing time decreases, differences become greater). This is found to be the case. The correlation ratio is -.96, and the regression is 1.15. The correlation ratio is highly significant.

Finally, since the hypothesis predicts that the means for grammatical and ungrammatical performance should diverge with a decrease in percentage playing time, we should expect a negative correlation between time compression ratio and the differences between the

averages of the spliced conditions (since these two lines are closest together, a strong correlation here would also mean a strong correlation between the differences between the means of the grammatical, normal prosody and ungrammatical spliced or normal prosody conditions). The observed correlation is -.59, and the regression is .188: moderately strong, although not in fact significant.

Discussion

The results of this experiment demonstrate clearly that the results of experiment three are not due simply to the "floor" effect as discussed previously. Experiment three extended the phenomenon observed in experiment two. Together, these three experiments show that the effect of an interaction between grammar and prosody is reliable and replicable. We may now turn to a discussion of the broader issues involved.

General Discussion

Firstly, with regard to Lieberman's "optimal coding" hypothesis discussed before, it is clear that some re-statement is necessary. Firstly, experiment one showed that a speaker will produce some prosodic cues to meaning even when he or she might be unaware of the fact that there is an ambiguity in what is to be said. Secondly, it has been demonstrated in experiments two to four that severe reduction in prosodic cues does hamper perception when the listener has to work in adverse conditions: correspondingly, we may assume that a speaker, being in some sense aware of this, will tend to exaggerate the prosodic contrasts in his speech in noisy and non-optimal communication situations.

From these results, we may amend the "optimal coding" hypothesis in two ways. Firstly, every speaker does in fact produce some prosodic contrasts when he speaks. In non-optimal situations, prosody may be one of the means whereby the speaker may be able to make his message intelligible. If there is any theoretical value in the "optimal coding" hypothesis at all, it is surely in the claim that a speaker relies more on context than on prosody to make himself understood. Such a claim may be difficult to substantiate or falsify, since prosody and context (of a non-prosodic kind) may well be giving different sorts of information, anyway.

To return to the results of the interaction, a statistical interaction between prosody and grammar was reported by Leonard (1973) in the context of a memory experiment. He found that intonational cues facilitated recall of "anomalous" (grammatical, but meaningless) sentences, but not of normal sentences, anagrams (ungrammatical, but meaningful) or word list strings. He concluded that intonation may serve as an additional component to grammar in the "anomalous" case. That is, intonational cues consistent with an implied grammatical structure facilitate recall.

There are two explanations which may be given for the interaction effects reported here and by Leonard. One explanation would give prosody a processing priority: that is, speech passes through a stage of prosodic analysis (specific to grammatical aspects) before it enters into the domain of grammatical analysis. The interaction is explained by saying that words strung together regardless of grammar have no properties that the grammatical component can relate to, and therefore, the results of prosodic analysis are largely wasted.

The other explanation is less specific as to processes, although it differs from the first in that it would explain the interaction in terms of prosody and grammar interacting rather like, to take a cybernetic analogy, a pair of "co-routines". This analogy refers to two programs, both of which pass control to each other at many stages in the computation, thus producing a complex "interactive" pattern of recursions. If the computer is denied access to either program, it cannot

proceed beyond a certain point with the computation. One may even posit a third explanation, which states that prosody is analysed after grammar: prosody thus extends its beneficial effects in grammatical sequences alone simply because since ungrammatical sequences do not get through the grammatical analyser, prosodic analysis cannot take place on them.

This last explanation is not very convincing because, one may well ask, what is the need for prosodic analysis on a signal which is already well-analysed by other means? It does not accord with introspective evidence, either. When listening to the tapes, for instance, of the ungrammatical normal prosody condition of experiment two, one has the feeling that the recording should be making sense; that sentences and phrases do exist, only somehow the perceptual apparatus is failing to deal with them. The effect is rather like that of listening to an incomprehensible lecture, enthusiastically delivered. It suggests that prosodic analysis should be considered as taking place before the grammatical: in the next chapter, this question will be discussed in greater detail.

Two papers appeared soon after the foregoing investigations had been completed and presented (Kirakowski and Myers, 1975) by Wingfield (1975) and Darwin (1975). They are of particular interest since they both discuss the notion of an interaction between prosody and grammar.

Both presented evidence from similar-looking experiments on the perception of "anomalous" sentences.

These sentences were constructed according to the methodology used previously by Garret, Bever and Fodor (1966) to demonstrate the importance of grammatical rather than intonational cues as primary for perceptual processing of sentences (but see also Wingfield and Klein (1971) who have cast doubt on some of Garret, Bever and Fodor's findings).

Normally, the major syntactic break (henceforth MSB) co-occurs with an intonational break or prosodic boundary signal (henceforth PB). These sentences were anomalous in that they were constructed by cross-splicing so that this normal regularity was violated: the PB could occur either before or after the MSB, producing a rather odd perceptual effect.

Wingfield (op. cit.) presented results from an experiment in which these anomalous sentences were presented at various levels of time-compression, from 100% to 50% in steps of 10%. Ss were instructed to listen to each sentence, and to write down what they had heard. Wingfield concluded that:

> the perceptual structuring of sentences is determined as much by prosodic features as by formal (grammatical?-J.K.) structure.

Furthermore, that

> perceptual segmentation is an intermediate stage of processing, dependent on an interaction of syntactic structure and prosodic cues.

He does not, however, present any data to demonstrate specifically that such an interaction is the case, and a model of additivity may well apply to his data--c.f. Levelt's comment on Wingfield's presentation (from Wingfield, op. cit.):

> your statement that the effects of syntactic structure and intonation are not simply additive... I see no evidence in your data that this is so. It should appear from, for instance, there being interaction effects in an analysis of variance.

Darwin (op. cit.) presented results from two experiments to support (in his terms) a "dynamic" theory of prosody in speech perception. Against such a theory he contrasted a model according to which prosodic information could be extracted to indicate syntactic structure independently of segmental information. His data on the different sorts of errors made by his Ss in the cases where the PB precedes the MSB and vice versa supports his notions of dynamism in perception. In his model, prosody guides the search towards an appropriate syntactic organization of the sentence "dynamically and interactively": in much the same way as syntactic and semantic constraints are used by Winograd's program (Winograd, 1972) of an integrated language understanding system.

Neither author mentioned above, however, states explicitly how such a "dynamic" or "interactive" system would combine prosodic and grammatical evidence, although Darwin does stress the importance of prosodic and grammatical prediction in ongoing perception. More importantly, although it seems to be accepted that prosody is attended to first to give a preliminary analysis, no evidence has been reported to demonstrate that it may not be the function of the segmental analysis to give this first "scan" through the input.

Wingfield does, however, point out a seeming paradox that if analysis for meaning follows initial segmentation, can initial segmentation be based on meaning? There may well be heuristics which can get around Wingfield's paradox: these questions will be treated in greater detail in the concluding chapter.

Since the structure of the interaction depends crucially on the nature of this preliminary analysis, the experiments to be reported in the next chapter will attempt to clear this point, by examining to what extent the first pass could be considered as prosodic. The chapter after that will examine how much prosodic information is actually conveyed by the speech signal.

Chapter Three

In this chapter, an attempt will be made to formulate a process model of speech perception with regard to prosody and grammar. The discussion so far has deliberately avoided the impedimenta of flow diagrams, and the interactive theory does not in itself imply any distinctions as to process.

The question to which this chapter will address itself to is: should the interaction between prosody and grammar be considered in terms of a model in which prosodic processing precedes grammatical processing, or a model in which prosodic and grammatical processing happen at the same time, mutually influencing each other? In terms of the first model, prosody could be considered in conjunction with other phonetic aspects of speech; in terms of the second, in conjunction with the grammatical.

General Introduction

The logogen model

A frame of reference is wanted that can in principle separate phonetic decoding from grammatical analysis. A strong set of claims in this regard have been made by Morton (e.g. Morton, 1970). His "logogen" model was evolved originally to account for phenomena of word identification independently of considerations of memory (Morton, 1964). It has subsequently been found to have relevance to memory phenomena (Morton, 1970), and again to speech perception (Morton and Chambers, 1976).

The heart of the model is the logogen system--a sort of neural dictionary which constantly adjusts thresholds for word recognition depending on the state of the rest of the model. For speech, input reaches it via the acoustic analysis system, and it outputs into a response buffer, a place where potential responses are stored. There is a feed-back loop between the logogen and the response buffer, equated in Morton's model with silent rehearsal. The logogen system interfaces with a cognitive system. In the cognitive system, a whole set of activities takes place to do with grammatical and semantic computations on input from the logogen.

Since input to the cognitive system is mediated (largely) through the activities of the logogen, we may assume that the primary activity of the cognitive system upon speech being heard is to produce some kind of grammatical analysis of the semantic and syntactic relationships between logogens. The acoustic system, on the other hand, will concern itself with phonological decoding, guided by a knowledge of the segmental and supra-segmental phonological system of the language, although it is the latter claim that will be investigated more fully later.

The entire system is shown diagrammatically in figure 3.1. An elaboration of the functions of the component parts mentioned above, and in particular an account of the importance of the visual component (which will not enter the present discussion at all) are more fully described by Morton (1970).

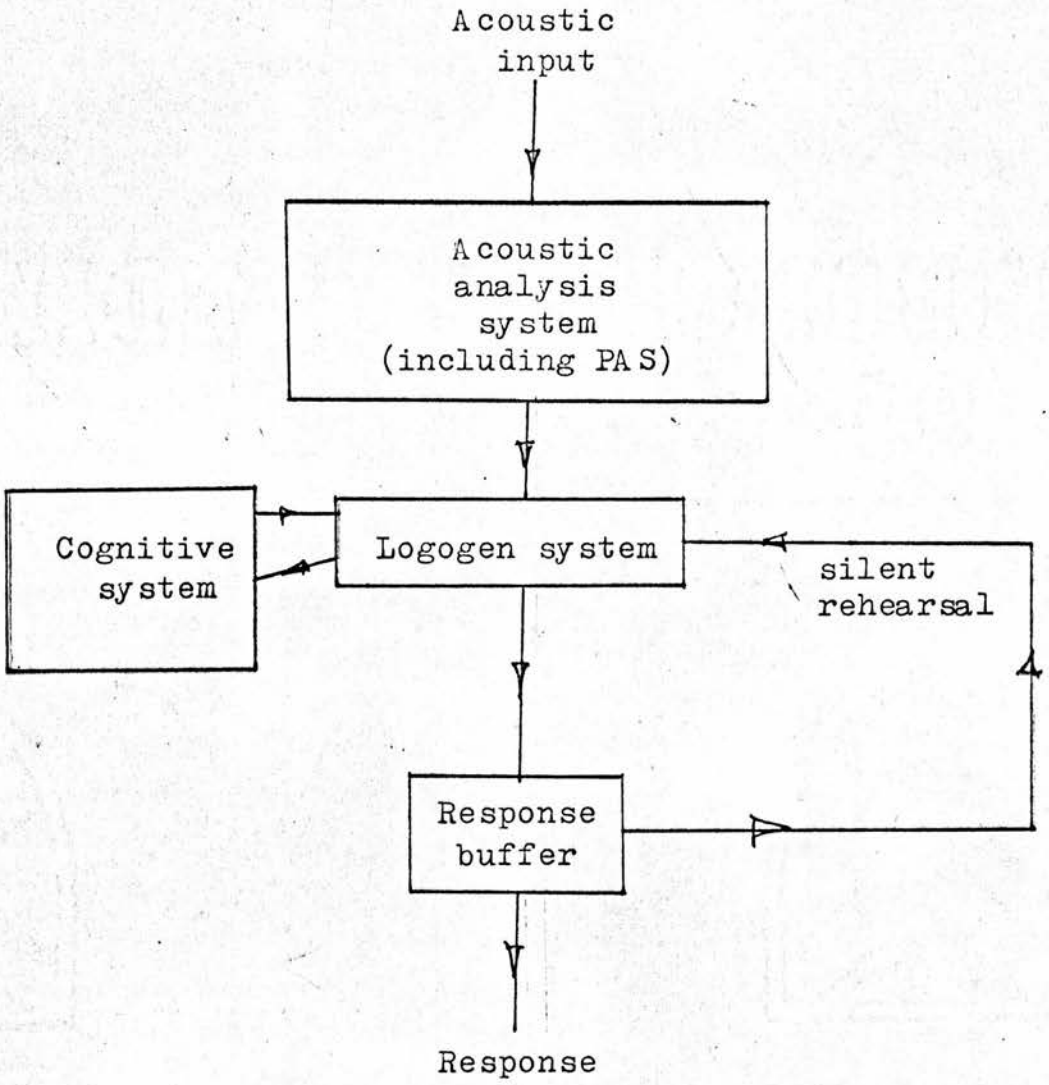

Fig. 3.1. A flow-diagram of information in the Logogen Model, after Morton (1970). The contribution of the visual component is not included here.

In terms of Morton's system as outlined above, the question is: is prosody attended to before or after the activities of the logogen system? If we can demonstrate that prosody is important before the logogen system's operations, that is, in the acoustic analysis system, then a case for prosodic pre-processing of speech is possible. On the other hand, if it is impossible to demonstrate such a phenomenon, then no real advance has been made: since all would happen in the cognitive system, Morton's model could not help us. This latter eventuality is a sort of null hypothesis for the moment, to which the experimental evidence will address itself.

The structure and function of the acoustic analysis system are the result of Morton's attempts to explain the "stimulus suffix effect" (hereafter simply "suffix effect") as parsimoniously but completely as possible in terms of the logogen model.

The effect is found in experiments on short-term memory for speech. If a spoken string of unrelated items is followed by a spoken suffix, recency (memory for the last few items of the string) is diminished in comparison with the case of a suffix which does not sound like speech. The effect was demonstrated by the following simple experiment (which was run as a demonstration of the suffix effect to the second year practical class in psychology).

Twenty strings of nine random digits were made up with replacement from the numbers 0 to 9. Half

the strings, randomly selected, were followed by a spoken suffix (the word "now") and the other half of the strings were followed by a non-speech suffix (a tone at 440 Hz. whose loudness and duration were equivalent to the average of these measurements on the speech suffix). The strings were recorded at the rate of two items per second. The suffix was recorded as if it were the tenth item of the string.

The task of the Ss was to listen to each string in turn, and to write the digits down in order of presentation as soon as they heard the suffix. The suffix itself was not to be written down. Ss were given about twelve seconds in which to make their responses, and each sequence was preceeded by an alert word ("ready").

Protocols were scored in strict serial fashion: that is, a digit was counted as a correct response only if it was the correct digit and reported at the correct serial position. Figure 3.2 displays the average probability of a correct response at each serial position for speech and tone suffix strings. The difference observed between the two conditions at the ninth serial position is significant beyond the 1% level of confidence when tested by Sandler's A statistic (computationally equivalent to student's "t"--see Runyon and Haber, 1968). That is, the ninth item is recalled significantly more often when it is followed by a tone suffix than when it is followed by a speech suffix.

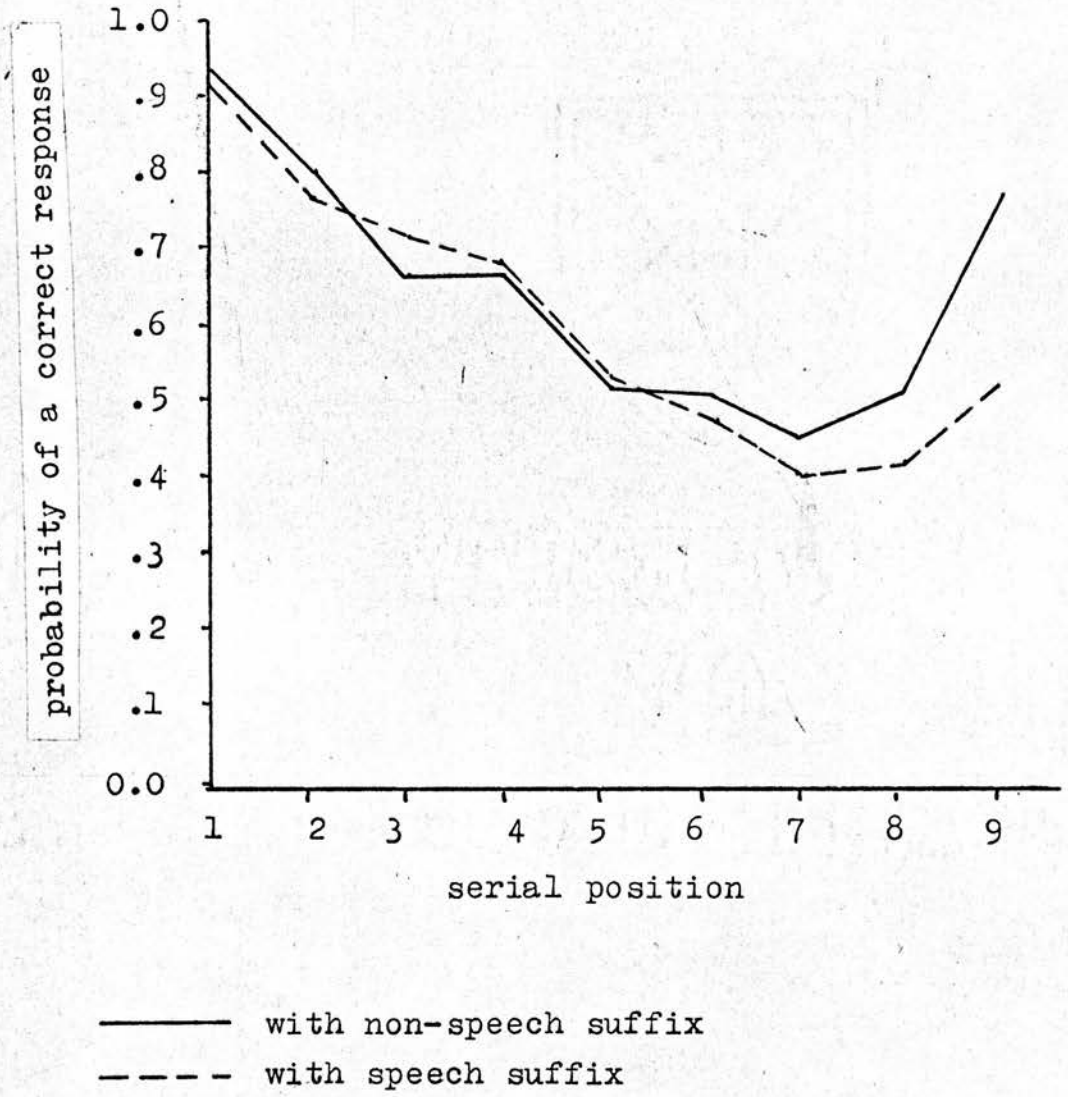

Fig. 3.2. Average probability of a correct response at each serial position for lists with speech and non-speech suffixes.

The methodological and procedural details of the above demonstration embody the classic requirements of a suffix experiment: random selection of the stimulus items, fast presentation of the sequence, the timing of presentation of the suffix, the instructions to the Ss, and the method of scoring the responses (although one frequently finds the results reported as probabilities of error where here we shall always report them as probabilities of a correct response).

The demonstration can be extended to show that the meaningfulness or loudness of the suffix does not contribute to the size of the effect (Morton, Crowder and Prussin, 1971). The timing of presentation, however, is important. If the items are presented slowly (fewer than one item per second) or the presentation of the suffix is delayed (by about two seconds) then the effect disappears or is severely attenuated (Crowder and Morton, 1969). The effect can be demonstrated with items other than digits, although it seems that some discrimination is made in favour of vowels against consonants (Crowder, 1972). The effect still holds if the Ss are given part or most of the presented sequence and are made to respond only with a few of the items or even just the final one, although it is not so great in this case (Morton, Crowder and Prussin, 1971). Scoring need not be strictly serial, although once again, such a procedure will show the effect at its strongest (Morton and Chambers, 1976).

A complete bibliography of papers relating to this effect has yet to be compiled, although some excellent partial reviews have been published (see for

instance Morton, 1970; Crowder, 1972). As the discussion below will show, however, the time is not yet right for such a complete review: much still remains to be discovered about the effect; and no doubt the theoretical interpretation will have to undergo consistent modifications.

The explanation hinges on the supposition firstly that there is a switch, which only permits speech-like sounds to enter the acoustic system. Readout from this system is buffered by a Precategorical Acoustic Store (henceforth, PAS) whose maximum capacity seems to be about three items, and which stores information for about two seconds. The last few items of the string are held therefore in PAS at the time of arrival of the suffix. Items presented earlier are sent, via the logogen system, to the response buffer to await output. From the time of presentation of the suffix, the logogen system is occupied with the organization of the response. If the suffix is sufficiently speech-like, hence admitted to PAS, it will interfere with the representation of the final items in PAS (which cannot enter the logogen system until the contents of the response buffer have been cleared by responding). Experiments with strings of three, five and seven items have shown that strings with five digits or less do not demonstrate a suffix effect (Kirakowski, in preparation). This suggests that what is held in PAS could be a backlog of information due either to the finite speed of operation of the logogen, or to the limited amount of storage space available in the response buffer.

Considerations of common sense also suggest that items arriving at the ears are immediately represented in some way in the cognitive system, as Morton (1970) claims. If this were not the case, responses would be impossible until the suffix has been processed from PAS. Morton, Crowder and Prussin also point out evidence to the effect that a large part of the suffix effect can be attributed to events occurring after some attention mechanism. The representation in PAS is obviously necessary in order to produce a correct response, given the terms of the explanation; registration in focal attention and the cognitive system may also be necessary, but not sufficient.

The acoustic analysis system should not, however, be considered solely as a passive receptor of information. It is quite in line with the logogen model to assume that what happens at PAS is a "phonological decoding of speech sounds" (c.f. Morton and Chambers, 1976). There is evidence to suggest that such a phonological decoding happens relatively fast, and is based on the past experience of the phonological decoder over a limited amount of time. Thus information from the phonological decoder could enter both the cognitive system directly, and be sent to PAS to await thorough processing by the logogen component (Kirakowski, Vance and Macnamee, in preparation).

The above explanation is summarised diagrammatically in figure 3.3. Once again, the account is not complete: for instance, the effects of monaural, binaural and mixed auditory presentations are not discussed since these are not germane to the issue at hand.

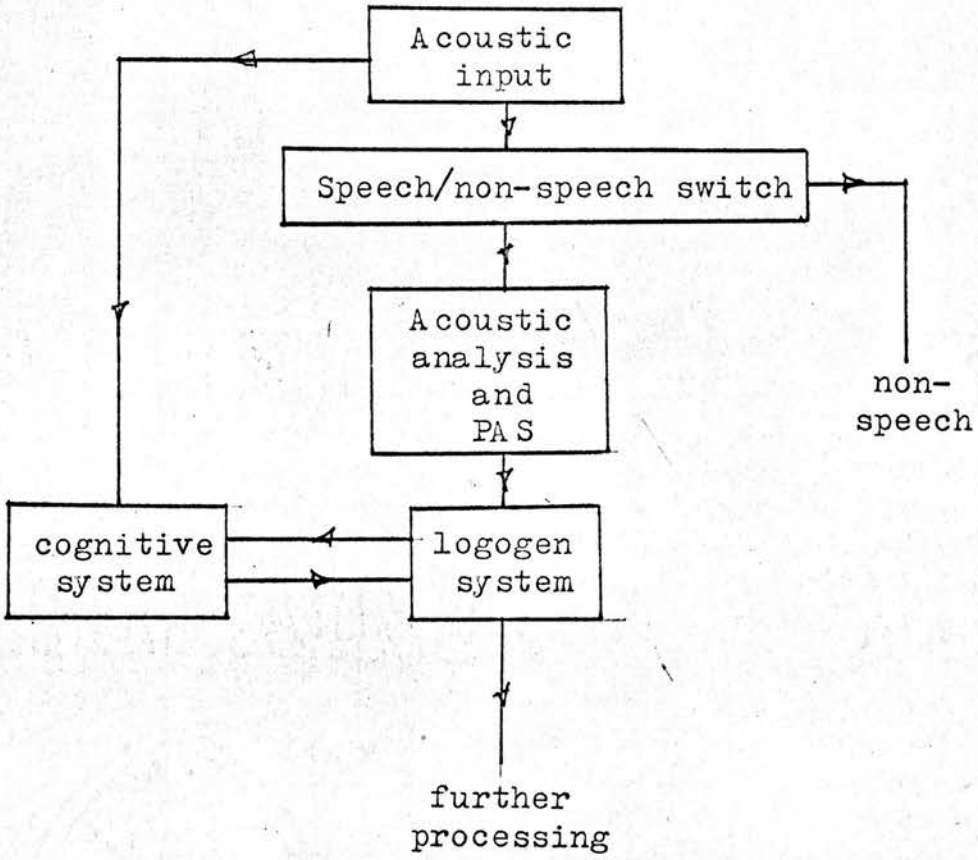

Fig. 3.3. Diagrammatic summary of Morton's explanation of the stimulus suffix effect.

PAS as a phonological decoder

Morton's explanation has been criticised by Massaro (1972) in a review of literature appertaining to what Massaro calls "pre-perceptual auditory memory". Massaro (see for instance Massaro, 1970, exp. V) reports that auditory similarity between a stimulus tone and a masking tone does not have any effect on the identification of the stimulus tone. However, some effects of auditory similarity are reported for the suffix effect: for instance, the effect is decreased if the suffix is read in a voice of recognisably different sex from that in which the rest of the list has been read (Morton and Holloway, 1970). From considerations such as these, Massaro concludes that the effect does not entail pre-categorical storage, but is due rather to interference of the S's short term memory for the final items in the list.

The force of this criticism is lessened when the different notions of "pre-categorical" held by Massaro and Morton are examined. Massaro's experiments are conducted using pure tones presented at fast rates for identification from a binary set (Ss have to identify a pure tone as either the "A" or "B" tone). Images in his pre-perceptual auditory store last for up to about 250 msec., whereas information in PAS can last up to two seconds. It may well be that the difference between these two stores reflects the fact that speech is more difficult to process than pure tones on account of its greater acoustic complexity, and consequently needs longer storage.

Morton and Chambers (1976) are quite prepared to believe that non-speech sounds are handled by a different system (after the speech-non-speech switch), and suggest Massaro's auditory store as a possible interpretation of such a system. Similar objections to the PAS explanation which refer simply to confusions in short term memory are rejected by Morton on the grounds that they fail to make the distinction between articulatory and acoustic effects.

More serious problems are posed by studies which, working in the framework of Morton's explanation, demonstrate a maintenance of recency despite a speech suffix. Salter (1975) manipulated the category of the final (pre-suffix) item: he used digit lists with a final letter and letter lists with a final digit. In a subsequent pair of experiments, (Salter, Springer and Bolton, 1976), the meaningfulness of the final pre-suffix item was manipulated: the lists were made up of paralogs of middle meaningfulness ("M") values, and the final paralog was either of low or high M value.

If the last items are represented in PAS only in their acoustic manifestation, these manipulations should have no influence on the size of the suffix effect. In fact, Salter found that the size of the suffix effect does fluctuate under the above conditions. A change of category on the last item over-rides the speech suffix and recency is maintained. The effect of high M paralogs before the speech suffix is to increase the probability of a correct response for the item to the level of the correct responses for a low M paralog when followed by a non-speech suffix.

Salter et al. discuss two interpretations of the phenomenon. Their first interpretation follows Morton's account fairly closely. Upon arrival, each item is sent automatically to the cognitive system, via the logogen system, and is also started on the PAS - logogen - response buffer route. The speech suffix induces confusion in PAS, to the detriment of the final items of the list. However, if any of these items are also strongly represented in the cognitive system, the probability of a correct response associated with that item increases. In the case of the alphanumeric experiments (i.e. Salter, , op. cit.), this representation is achieved by analogy with the von Restorff effect; in the case of the paralog experiments, it is achieved by the differential semantic codability of the items (the von Restorff effect is that if a list of items is presented in a short term memory experiment, and one of the items stands out by virtue of being perceptually salient, this item will be remembered better than other items in the list). In either case an additional postulate is necessary to the effect that (semantically) similar items arriving in rapid sucession at the cognitive system are confusable. Considerations of the general mode of function of memory would then predict that earlier items would show more confusions than later items (Morton, 1970, p. 246). Morton does not regard this experiment as having been carried out, although if the explanation of Salter et al. is to be regarded as generally true, it would have to hold for, among other things, strings of digits, which are eminently confusable. As has been demonstrated, the opposite is the case. Items arriving early in the

sequence are reported better than items arriving later (see figure 3.2).

A second interpretation which is embodied in two "proposals" diverges more radically from Morton's explanation. First of all, it locates the speech-non-speech switch within the acoustic system, and argues that it is not specifically attuned to speech alone. Secondly, it posits that all incoming items activate representations of themselves in the cognitive system without necessarily having to pass through a limited capacity processor. Confusions in the cognitive system arise if other (presumably acoustically similar--c.f. Morton and Chambers 1976) items follow rapidly.

This second interpretation, so radically at variance with what has been discovered about the components involved, is best considered as an alternative explanation rather than a modification of the original one. It remains to be seen whether it can explain the known facts about the suffix effect and the cognitive system better than the original explanation, and whether it can generate any new, testable, hypotheses about the suffix and allied effects. The first interpretation, therefore, is to be preferred at present.

Kirakowski, Vance and Macnamee (op. cit.) investigated this first interpretation. They contrasted it to one in which PAS or some property of the acoustic analysis system was responsible for the increase in recency in the Salter demonstrations.

If it were the cognitive system, then any mistakes made in reporting a pre-speech-suffix item would reflect the predominant mode of activity within the cognitive system: that is, they would be semantic mistakes (for instance a high proportion of synonyms or antonyms would be expected). If recency were ascribable to the acoustic analysis system, the preponderance of errors would be acoustic ones.

The experiment was carried out using CVC syllables of high association value for the lists. A random half of the lists has a meaningful monosyllabic word as the final item; the other half had another CVC syllable. All lists were followed by a speech suffix.

First of all, the result of Salter et al. was replicated. More meaningful words were recalled than CVC syllables in the pre-suffix position. An analysis of Ss' errors on the meaningful words revealed that of a total of 89 errors made in this position (out of a possible total of 220 words), 35 were classifiable as acoustic errors; and two as acoustic and semantic errors--they were both homonyms, from two different Ss. The remaining 52 errors were no response (which is treated as an error). Twelve of the acoustic errors were meaningful words which bore no semantic relationship to the stimulus word. Thus whatever system was implicated in these responses, it was certainly not one which made semantic confusions.

This evidence taken as a whole suggests that the acoustic analysis system is responsible for the observed recency effect. A model analagous to one proposed by

Figueroa (1978) for visual iconic storage was suggested. Since meaningfulness and association value are correlated with the degree of conformity to phonological structure, PAS could simply be pre-disposed to favour items which displayed some degree of phonetic similarity to English (or whatever language the experiment happens to be conducted in). The acoustic analyser should also be able to maintain information about the phonetic environment of an item over at least the length of a tone group (c.f. Cutler, 1976) and it might be expected that PAS treats phonetically novel items differentially. That is, one basis for determining how important an item is and therefore its length of stay in PAS could be the frequency of occurrence in the stream of speech immediately before that item.

To test the hypothesis of frequency of occurence, Kirakowski, Vance and Macnamee carried out a suffix experiment in which lists of digits were presented in groups. Within each group, one digit chosen at random never appeared till the final pre-speech-suffix position of the last list of the group. Apart from this, the distribution of speech and non-speech suffix conditions was random. Ss had to report on each list as usual, and the grouping of lists was not made apparent. It was found that withholding and presenting items in this way increased their recency in the speech suffix condition, when compared to items which had not been treated in this fashion. The findings of this experiment confirm the above interpretation of the activity of the acoustic analyser.

Some space has been devoted to this issue since it introduces the concept of PAS as an active phonological decoder, sensitive to some contextual effects; and also because an interpretation of the following two experiments à la Salter et al. might weaken the import of the conclusions to be drawn from them to the matter of priorities of processing.

Experiment 5

The procedure in preparing stimuli for an auditory short term memory experiment is usually to have them read evenly, on a monotone and at equal subjective intensities. In order to investigate prosodic effects, this practice will be violated. The principal experimental hypothesis is that a stressed digit in a pre-speech-suffix position of a nine-digit list will have greater recency than an unstressed digit in the same position. In addition, the effect of stressed digits in serial positions 3, 5, and 7 will be investigated: the hypothesis here is that stressed digits at these positions would also be recalled better than unstressed digits in the same position. Stresses in these positions are included since Ss might develop a "set" on the ninth position if this were the only one which ever received a stressed digit, although the suffix effect seems to be relatively independent of behavioural strategies (Crowder, 1969).

Morton, Crowder and Prussin (1969) have manipulated the intensity of the speech suffix. They found that suffixes twice as intense (subjectively) as the average list item had less effect than suffixes of the same intensity. They suggested that at the point in the system at which the effect takes place, intensity does not have an analogue representation (i.e. they argued that it is not the case that the greater is the intensity, the stronger or larger the representation). Instead, they argued that it is represented neutrally--one might say that intensity

is represented as a feature. Thus a selector could use the intensity feature cue as a basis for discriminating between items. This interpretation is consistent with the concept of the acoustic analysis system as an active phonological decoder--as discussed above.

Since intensity is itself one of the acoustic cues to stress (see Lehiste, 1970, chapter 4) a reasonable hypothesis to make in the circumstances could be that incoming items are not only phonetically identified by the acoustic analysis system, but also characterised in terms of stress. It is therefore interesting to examine whether the cognitive system would use such stress information with regard to the organization of recall of the earlier part of the list. From a consideration of the mode of function of the cognitive system, one would expect the feature of stress to be treated as a semantic attribute of such an item (in the sense that Halliday considers the tonic to be the informational focus of the tone group).

Method

Design. Twenty random nine-digit lists were generated from the numbers 0 to 9 with replacement. Six sequences were assigned at random to condition four, five each to conditions three and two, and four to condition one. The order of sequences was random. In condition one, the third digit of the list was to be stressed, in conditions two, three and four, the fifth, seventh and ninth respectively. A speech suffix followed

each list directly. Ss' instructions were to listen to each sequence, and upon hearing the suffix, to write down the digits in the order in which they had been presented. If Ss were unable to recall a digit they were asked to guess or put a short dash. Protocols were to be scored in strict serial order (henceforth the "serial" method of scoring): that is, a response was considered correct only if it consisted of the correct digit in the right place. Dashes indicating no response were scored as incorrect responses.

Procedure. The digits were recorded on tape and presented to Ss on headphones in "parallel tracking"-- that is, the same signal reached both ears simultaneously. The instructions to the Ss were on the tape. These were followed by seven practice runs, in which an example from each condition was presented, and then the experimental trials. There was a short break after the first ten trials.

Each trial proceeded as follows. There was an alert word "ready", followed by one and a half seconds of silence, then the list, at the rate of two items per second, followed by the suffix word "now" spoken in tempo as the tenth item of the sequence. Eleven seconds of pause then ensued to allow the Ss time to recall before the next trial began.

The unstressed digits and the suffix were spoken on a monotone at equal subjective intensities. Stressed digits were characterised by a rise and fall in intonation

on the voiced part, and a subjective increase in intensity. The digit "0" was pronounced "Oh" after some enquiries as to local usage.

Timing was effected by E reading in time to metronome clicks recorded on a tape loop that was played back through headphones. These clicks were not audible on the stimulus tape.

Ss made their responses in rows on lined sheets of paper. They were instructed not to write down the alert word or the suffix, not to start responding till they heard the suffix, and to recall the digits in the order in which they were presented. No constraint was imposed on the number of items the S could write down. Sometimes Ss recalled less than nine, and sometimes as many as ten or eleven.

Subjects. These were 36 psychology students.

Results

For each S, the probability of a correct response was calculated at each position for each of the four conditions. Figure 3.4 displays the average probabilities thus obtained. Table 3.1 displays the obtained total probabilities, unaveraged. This data was the basis for all subsequent analysis.

To avoid confounding serial position effects with the dependent variables of the experiment, a separate analysis of variance was computed comparing

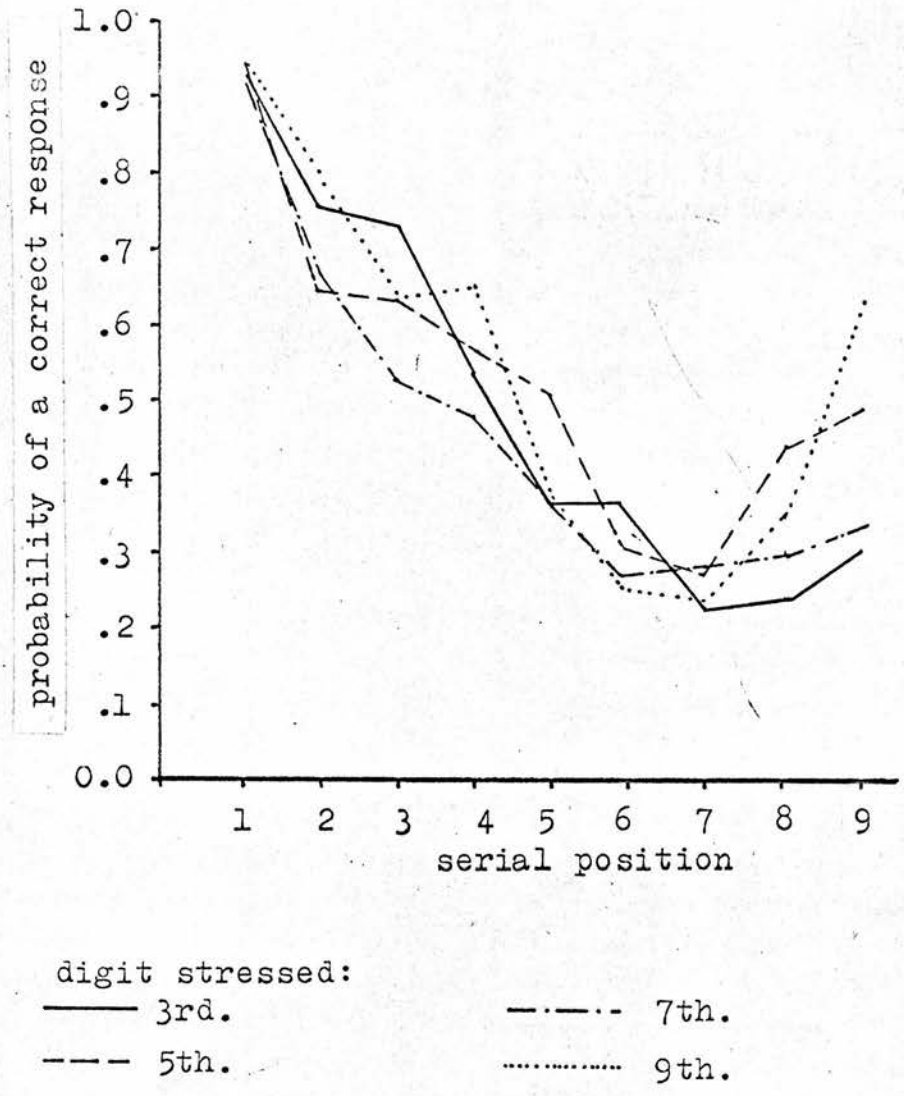

Fig. 3.4. Average probability of a correct response at each serial position for lists with 3rd., 5th., 7th., and 9th. digits stressed. "Serial" method of scoring.

TABLE 3.1: Total probabilities of a correct response in each condition at each serial position, totalled over all Ss in experiment five.

stress at position:	3	5	7	9	MS res.
Serial position: 1	33.2	34.0	33.0	34.3	.0117
2	27.3	23.6	24.2	28.8	.0286
3	26.5	23.0	19.6	23.0	.0367
4	20.0	20.6	17.6	23.8	.0461
5	13.7	18.6	13.8	14.8	.0507
6	13.5	11.4	10.0	9.8	.0386
7	8.7	10.4	10.6	9.0	.0258
8	9.0	16.2	11.6	13.0	.0358
9	11.5	18.4	12.6	24.3	.0415

Total number of Ss = 36. The figure for MS residuals is included here from the analyses of variance.

obtained totals in the four conditions for each position. The experimental hypothesis predicted that positions three, five, seven and nine should show a significant effect, while any variation observed between conditions at other serial positions should be ascribable to chance. The statistical model was a single-factor repeated measures analysis of variance (Winer, 1974). The results of the analysis are shown in table 3.2. The null hypothesis was rejected at positions three, five and nine. At position seven, there was no effect of conditions, and the hypothesis predicted that condition three would have superior performance. A priori tests for differences between totals were only carried out, therefore, at the significant positions. Table 3.3 shows the results of these tests. The null hypothesis is rejected every time at the 1% level of confidence.

TABLE 3.2: Results of the nine analyses of variance on the results for the four conditions of experiment five at the nine serial positions.

Position in list	Obtained F	Significance
1	.928	NS
2	6.032	.01
3	6.010	.01
4	3.981	.01
5	2.878	.05
6	2.074	NS
7	.971	NS
8	6.987	.01
9	23.428	.01

d.f. = 3,35.

TABLE 3.3: *A priori* tests for experiment five.

Position in list	C^2	F	Significance
3	13.9^2	12.186	.01
5	13.416^2	8.218	.01
7	-	-	NS
9	30.493^2	51.864	.01

However, the unexpected significant main effects at positions two, four and eight also require examination, and since there was no specific hypothesis, the Neuman-Keuls procedure for *a posteriori* analysis of the significance of differences between totals was used (following Winer's method; Winer, *op. cit.*). The results of this procedure applied to all positions which were significant from the analysis of variance are summarised in table 3.4.

TABLE 3.4: *A posteriori* tests on experiment five.

Position in list	Summary of significant order of conditions
1	NS
2	4>(2,3)
3	1>3
4	4>3
5	NS
6	NS
7	NS
8	2>1
9	4>2>(1,3)

All significances calculated by the Neuman-Keuls method with $p<.01$.

It will be seen, that in four out of the five comparisons made, condition three has one of the lowest if not the lowest total, and one may speculate whether the insignificant result obtained at the seventh serial position is not in fact a reflection on Ss' generally poor ability to cope with condition three.

Some theoretical reasons why condition three could be more difficult to cope with will be discussed at the end of this experiment; at the moment we may entertain the a posteriori hypothesis that this result was due not to Ss' inability to recall the items at all, but to their inability to recall them in their correct order. If this were the case, a more generous method of scoring the protocols that does not place such a high penalty on the incorrect reporting of one digit's serial position should reveal a significant effect at position seven.

The method used for re-scoring was to count a response as correct not only if a S reported the digit in question at the correct serial position, but also if he reported it to one position on either side. Once a digit in a protocol has been counted as a correct response for a given serial position, it cannot be entertained as a candidate for any other serial position. This is the "one-either-way" method of scoring (see table 3.5). It is a compromise between promiscuously scoring correct responses (i.e. regardless of serial position) and the strict "serial" method used previously. Since the probability of all nine of the digits appearing within any one list is fairly

TABLE 3.5: Total probabilities of a correct response in each condition at each serial position, totalled over all Ss in experiment five. "One-either-way" method of scoring.

stress at position:	3	5	7	9	MS res.
serial position: 1	33.7	34.2	34.2	34.7	.0094
2	31.3	31.0	31.6	30.7	.0231
3	34.0	30.8	26.2	27.5	.0259
4	29.0	26.6	25.0	28.5	.0405
5	21.7	26.8	24.2	21.5	.0529
6	20.5	19.2	15.6	20.8	.0401
7	14.2	18.0	18.2	16.0	.0334
8	17.0	27.0	20.4	18.3	.0441
9	14.5	20.8	17.2	28.5	.0389

Total number of Ss = 36. The figure for MS residuals is included here from the analyses of variance.

high, such a promiscuous method would tend to blur any observable effects.

The one-either-way method does raise problems of comparison between scores as serial positions one and nine (there being no "zeroeth" position responses by definition, and few tenth position responses), and scores at serial positions two to eight. In fact, comparison of the totals between tables 3.1 and 3.5 shows that at positions one and nine the difference between the two methods is less than at positions two to eight. Since no comparisons are made across serial positions, this is not such an important consideration, although it should be borne in mind when examining figure 3.5, which displays the average probability of a correct response in all four conditions and nine serial positions. No doubt a correction could be made by dividing the scores at positions one and nine by two and those at positions two to eight by three to reflect the number of positions over which a response was sampled; but this manipulation may take us further from the real outcomes rather than nearer to them.

Table 3.6 summarises the results of the main effects analysis of variance (the same statistical model as before was used). In general, this method has lessened the differences between totals in two positions where no differences were expected (viz. positions two and four); magnified the differences in the case of serial position seven, but also magnified the differences in the case of serial position six.

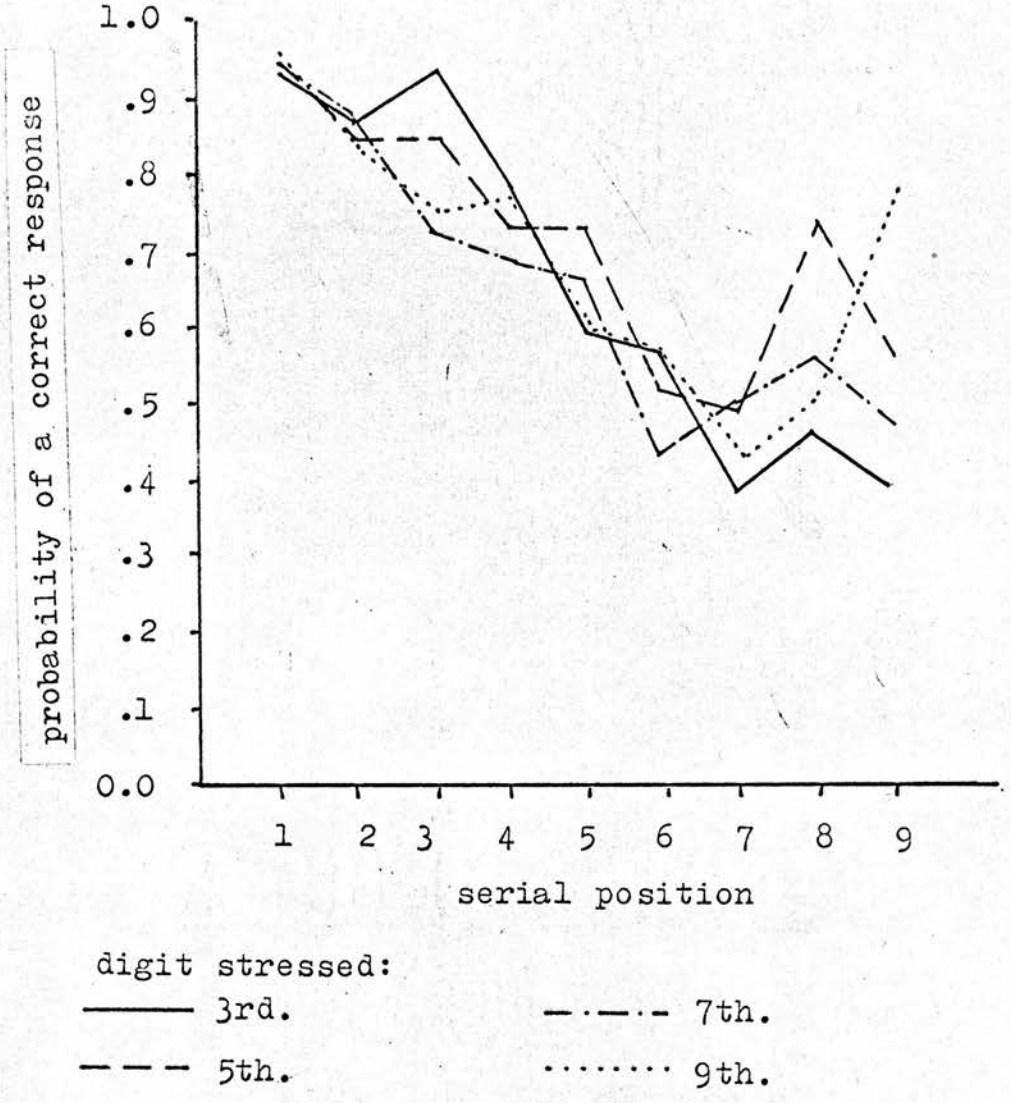

Fig. 3.5. Average probability of a correct response at each serial position for lists with 3rd., 5th., 7th., and 9th. digits stressed. "One-either-way" method of scoring.

TABLE 3.6: Analyses of variance for the data of experiment five scored in the one-either-way method.

Position in list	Obtained *F*	Significance
1	.419	NS
2	.194	NS
3	13.158	.01
4	2.316	NS
5	3.228	.05
6	3.976	.05
7	2.892	.05
8	12.402	.01
9	26.348	.01

d.f. = 3,35.

A priori tests disclosed the same pattern as previously: the totals for conditions one, two and four are significantly higher than the rest at the predicted positions, but condition three, although significant in the overall analysis, still fails to reject the null hypothesis. The obtained *F* value (2.793) is near the required value (d.f. = 1,120; *F* = 3.94 at $p<.05$).

Of the *a posteriori* tests carried out on all the positions with a significant *F* value in the main analysis, only three are significant out of the six, namely at positions three, eight and nine. The analysis suggests, once again, that poor performance in condition three could be an important factor, but also that some of the significances observed could be ascribed to better than average performance in condition two (see table 3.7).

TABLE 3.7: A posteriori tests on the data from experiment five scored the one-either-way method.

Position in list	Summary of significant order of conditions
1	NS
2	NS
3	1>(4,3) and 2>3
4	NS
5	NS
6	NS
7	NS
8	2>(1,3,4)
9	4>2>(1,3)

All significances calculated by the Neuman-Keuls method with $p<.01$.

The statistical analysis of the re-scored data does not suggest that the only reason why the totals at position seven do not differ is because the original method of scoring paid too high a premium on reporting at the correct serial position. It is therefore more than likely that the decrement in performance observed in condition three is due to interference not only with serial order information but also with information about the items themselves. Why this should be so is puzzling, and will be discussed again later.

Summary of conclusions

The experimental hypothesis, predicting that a stressed digit in the list final position would survive the suffix effect better than an

unstressed digit in the same position, has been proven. Stressed digits in positions three and five were also recalled better than their unstressed counterparts. Stressed digits in position seven did not appear to be remembered better than unstressed digits at the same position. A posteriori tests disclosed that a list with a stress in this position appeared to be harder to recall overall than lists with stresses elsewhere. The data was re-scored to investigate whether this result was due to a strict serial method of scoring, and it was discovered that it could only be so in part.

Discussion

A more general discussion of the importance of the findings in relation to a theory of speech perception will be found at the end of this chapter, when results from both this and the next experiment will be discussed together. For the moment, an issue specific to our experiment has to be dealt with: that of performance at serial position seven.

The original explanation of the suffix effect divided up processing of the stimulus list between PAS and the response buffer at the time of arrival of the suffix. Of a nine-digit list, the response buffer would hold six or seven items and PAS two or three. From this, it is clear that the seventh item occupies an ambiguous position: it is sometimes represented in PAS and sometimes in the response buffer.

An examination of the lines of performance over all positions for conditions one, two and four discloses that the probability of a correct response is at its lowest at position seven. This could well be a general feature of suffix experiments with nine items (see figure 3.2, for instance).

If the seventh position is occupied by an item which has been recognised by the acoustic analyser to be of some importance (e.g. it was recognised as being stressed), the system could well be in some quandary when the logogen, at the time of arrival of the suffix, refuses to accept any more input from PAS. The attempts of the acoustic analyser to pass the all-important item along could well interfere with the operation of the logogen in preparing the response as well as being an additional source of interference to that from the speech suffix, which latter is by then residing in PAS.

The data from the entire experiment was subjected to a three-way analysis of variance (positions by conditions by subjects), using the three-way interaction term as an estimate of error (Winer, 1971). The effect of conditions overall was significant, and the Neuman-Keuls *a posteriori* analysis disclosed condition three to yield significantly lower response probabilities than the other three, on both methods of scoring. The results of these calculations are included in the appendix.

Experiment 6

In this experiment, the effect of introducing some perceptual coherency to the stimulus list by means of intonational contours will be examined. If the acoustic analyser is a system which handles intonation, signalling the end of a digit list with an intonational device and then presenting a suffix on another intonation contour should produce a coherency to the digit list which in a monotonically presented sequence is lacking. This coherency would therefore manifest itself in an increase in recency for the last few items of the list.

A superficially similar experiment has been carried out and reported by Mills and Martin (1974). Their study was concerned with the prefix effect. The prefix effect is also found in short term memory experiments for acoustically presented materials, under much the same circumstances as the suffix effect. Only that instead of a redundant suffix being presented after the list, a redundant prefix is presented before the stimulus list. The effect is to depress the average probability of a correct recall for every item of the list when compared to a condition with no prefix (Dallett, 1964). The prefix effect does not admit to the same sort of explanation as the suffix effect: as has been demonstrated the speech suffix affects only the last few items, whereas the prefix affects recall for the entire list.

Mills and Martin tested the effect of what they called "articulatory organization" . Martin (1972) discussed the concept of "rhythmic patterns": among the properties of such a rhythmic pattern is one that enables a part of the pattern to potentially contain cues or information concerning the preceding and following parts. This rhythmic pattern is said to correspond acoustically to an articulatory gesture. The hypothesis they tested was that the probability of recall of items when the prefix and the list were not part of the same gesture would be higher than when they were. In their experiment, however, the stimuli were read evenly at the rate of two items per second in a steady monotone. It is difficult to see what precise acoustic cue gave listeners the impression of articulatory intactness, and Mills and Martin do not elaborate on this. The theoretical explanation of the prefix effect is also somewhat obscure, although a lot of evidence has been gathered about it (see Crowder, 1967).

In our experiment, the experimental condition will present the list of digit stimuli on a separate intonation contour from the suffix. In order to keep the manipulations as nearly the same as possible, in the control condition, the list and suffix will be presented in one intonation contour.

Method

Design. Twenty random nine-digit sequences were made up as for experiment five. Ten sequences were assigned at random to condition A, and ten to condition B. In the experimental condition (A) the list was read in one tone-group, followed by the suffix on a one-item tone group. In the control condition (B), both sequence and suffix were read in one complete tone group. The rest of the design was exactly the same as for the previous experiment.

Procedure. With the exception of the method of recording the stimuli, the procedure for this experiment was exactly the same as for the last experiment.

Each tone group was read on a slightly rising - falling contour (in Halliday's system, a tone of ...1+, with a listing pretonic), unless it was the suffix on its own when it was a sharply falling contour (tone 1, no pretonic). The contour's peak, or tonic, was on the seventh digit in the experimental condition (two tone groups per sequence) and on the eighth in the control. The reason why the tonic digit was not always placed at the end was because the purpose of the experiment was not to replicate the findings of experiment five which had already tested the effect of a final stressed digit.

A listening test was carried out on three Ss. They were supplied with a copy of the digits recorded for each list. The tape was played to them and their task was to read the sequences silently to themselves during presentation, and then underline which (if any) digit or digits were made prominent by the way the list was read. They were instructed to ignore the alert and the suffix. They were also allowed not to underline any of the digits of a sequence if they thought all the items were equally prominent. The results show that the only significant differences in prominence between the experimental and control conditions was that in the experimental condition digit seven was more prominent and that in the control, digit eight was more prominent. This difference was significant at the 1% level of confidence.

Subjects. Ss were 39 undergraduate psychology students, none of whom had done the previous experiment.

Results

Results were scored both serially and by the one-either-way method. Figure 3.6 displays the results for the serial method, figure 3.7 for the one-either-way method.

To avoid confounding serial position effects with the dependent variables of the experiment, a separate test for significance of differences between means was carried out on each pair of means at each

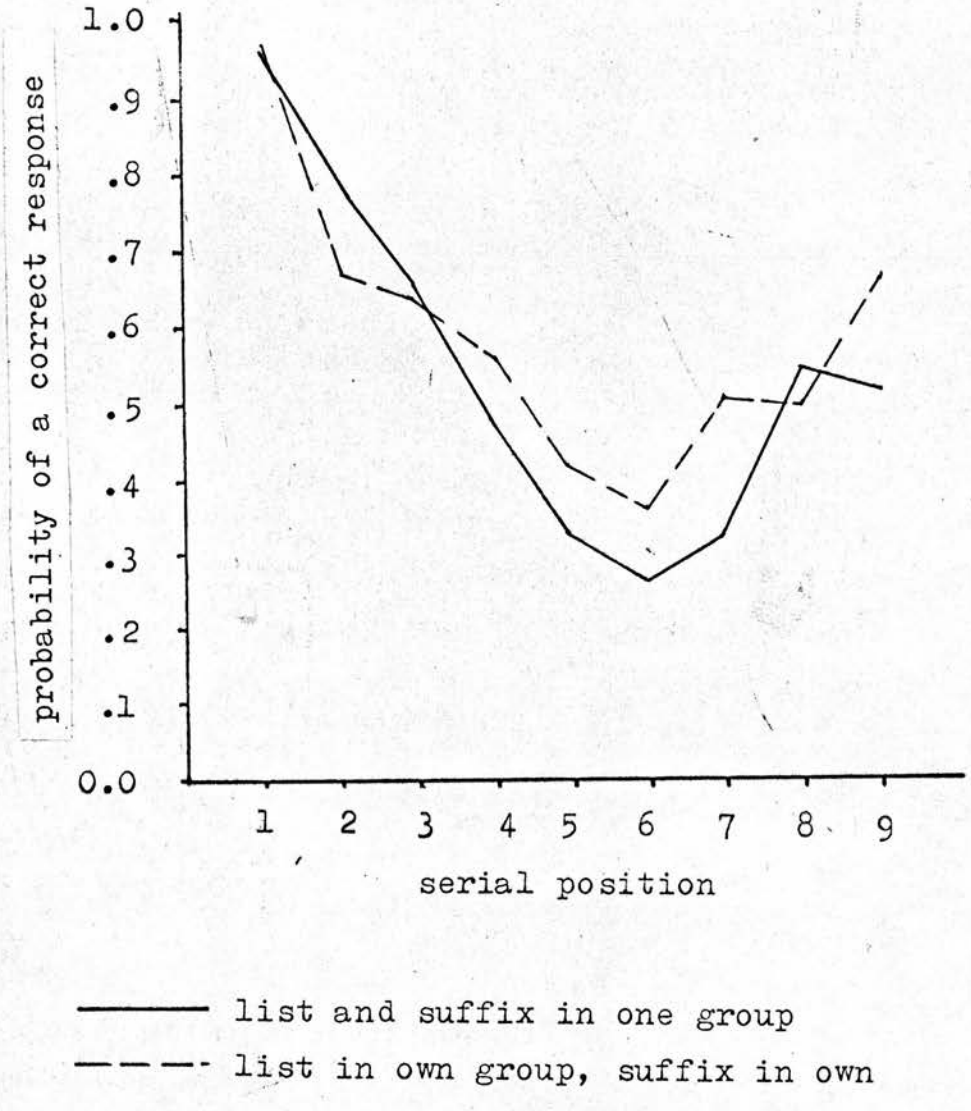

Fig. 3.6. Average probability of a correct response at each serial position for lists and suffixes read in one tone group, and lists and suffixes read in different tone groups. "Serial" method of scoring.

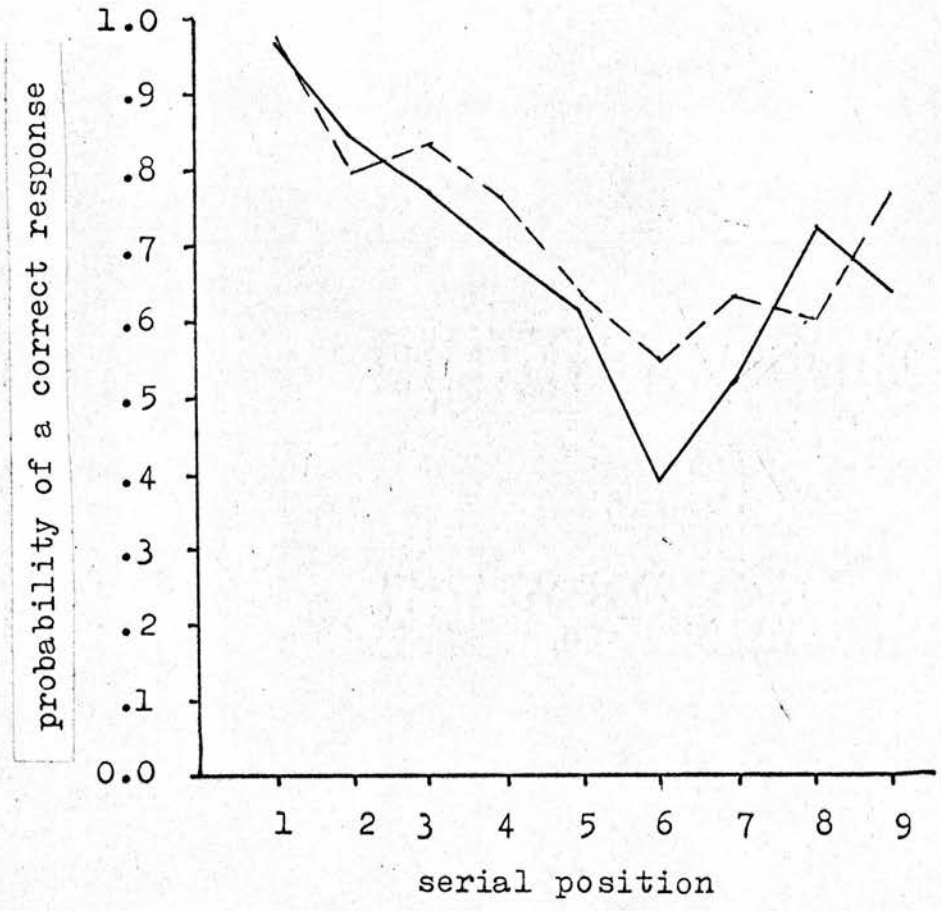

Fig. 3.7. Average probability of a correct response at each serial position for lists and suffixes read in one tone group, and lists and suffixes read in different tone groups. "One-either-way" method of scoring.

serial position. The experimental hypothesis predicted that the only position which should show any effects was position nine, and possibly also positions seven and eight, although the effects of the intonation contour may produce artefacts in this latter regard. The test used was Sandler's A statistic. The results of these computations are shown in tables 3.8a and 3.8b.

First of all, the differences between the means at position nine were as predicted by the experimental hypothesis in both methods of analysis. The results for position seven give a recall advantage to the experimental condition, although this is not statistically significant in the one-either-way method. As for position eight, in both methods, the mean for the control condition is higher than the mean for the experimental condition, although this is not statistically significant in the serial method. As mentioned above, these results may reflect on the way the stimuli were recorded rather than on any direct properties of PAS. The significance of the results for position seven is not directly relevant to the discussion of experiment five's results at this position, since the (proven) perceptual coherency of the list of digits in our experiment may well be a confounding factor in terms of the explanation given.

TABLE 3.8: Average probabilities of a correct response at each serial position for both conditions in experiment six tested by Sandler's A statistic.

(a) Serial method of scoring.

Position	Experimental Condition	Control Condition	A	Significance
1	.981	.969	1.50	NS *
2	.675	.806	0.12	.01 *
3	.637	.662	5.75	NS *
4	.569	.487	0.37	NS *
5	.419	.337	0.44	NS *
6	.356	.269	0.41	NS *
7	.512	.331	0.19	.05 *
8	.512	.550	1.00	NS *
9	.669	.519	0.16	.01 †

(b) One-either-way method of scoring.

Position	Experimental Condition	Control Condition	A	Significance
1	.981	.981	1.00	NS *
2	.806	.856	0.34	NS *
3	.837	.781	0.88	NS *
4	.775	.694	0.29	NS *
5	.637	.606	2.80	NS *
6	.562	.387	0.13	.01 *
7	.637	.544	0.32	NS *
8	.606	.731	0.20	.05 *
9	.769	.637	0.13	.01 †

d.f. = 38; $p < .05$.

* two-tailed test
† one-tailed test

General Discussion

Whatever the difficulties encountered in experiment five with the stressed seventh digit, it is clear that support for the general hypothesis has emerged strongly from the presented data. That is, we have good grounds for supposing that attention to prosodic features of the speech input is paid when the input is at a pre-categorical stage--before any lexical identification has taken place and therefore before any grammatical operations.

Somewhat more sceptically, it may be argued that going from the result of an experiment on the recall of random digits to theories of speech perception is a bold theoretical step to take. The essence of Morton's model, however, represents the essential components of any speech perception system and the use of short term memory paradigms is simply a device whereby the workings of the acoustic analyser might be further investigated. No less an authority than George Miller, after all, is reported to have said:

> that he was not dismayed by the number of memories or storage buffers hypothesised by students of information processing. A storage buffer is like a mailbox, and he would expect the nervous system to be full of buffers, just as the country is full of mailboxes (quoted in Kavanagh and Mattingly, 1972, p.289).

Experiment five showed that a stress on a final item gives it a recency advantage. This is

comparable to the effect of manipulating the alpha-numeric category of that item, as shown by Salter (1976). Experiment six showed that introducing an intonational contour to the digit list also had a positive effect on the recency of the final (pre-suffix) item so long as the suffix did not share the same intonational contour. This effect is separable from that shown in experiment five, in that, in experiment six, the stress of the final item was not a controlling variable. There are therefore two separate sources of information which give rise to the enhanced recency of the final item: contour and stress.

There are at least two possible explanations for this recency effect. One states that PAS simply maintains important items in PAS store, and as a result of the extra effort taken by PAS to maintain this information such items are less liable to be confused when the speech suffix arrives. The other states that the acoustic analyser recognises the importance of such an item, as before, but rather than maintain it in the relatively impermanent PAS system sends it on to the cognitive system, whence it is retrieved specially when it is necessary for recall. Introspective evidence seems to suggest the first possibility: if the ninth digit has been stressed, one can hear it ringing in one's ears for a while even after it has been written down. The second explanation too, would need elaboration with respect to how the item can be reported in the correct serial order, if it has been treated

differently from its context in the sense of being placed in another storage location.

Considerations of the results of experiment six would also suggest the first explanation to be more likely. Here, the second would have to posit that the entire list is somehow represented in the cognitive system upon termination of the intonation contour. If this were the case, one might expect that lists recalled from the cognitive system in their entirety would give a different shape to the recall curve than lists recalled partly from there and partly from PAS. Unfortunately, in the present experiment, the only digits which could give us some evidence about what was happening by deviating from the normal suffix-effect pattern were confounded with the independent variable (viz. contour peak on digits seven and eight). Finally, this explanation would have to square with the results of the experiments by Kirakowski, Vance and Macnamee, discussed earlier.

What is clear is that whatever else may take place in the cognitive system at the time of presentation of the suffix, the discriminations noticed in the presented experiments are motivated by phonetic rather than semantic criteria. The evidence also supports the notion of an active acoustic analyser, although how much processing is done by the analyser and how much by the logogen system is not so clear.

Morton, Crowder and Prussin's experiments in which the intensity of the speech suffix was manipulated led them to conclude, as mentioned before, that intensity was not coded in an analogue way: i.e. it was not the case that the greater the intensity, the more extensive the message. The most parsimonious positive suggestion from this conclusion is that the logogen system receives information in the form of a bundle of phonetic features--perhaps even the distinctive feature matrix? Such a bundle would include information such as stress and place in intonational contour, not traditionally part of the account of the distinctive feature matrix, so this identification must needs remain highly speculative.

However, this gloss in itself is no longer enough to satisfy the evidence. On sheer probabilistic grounds, providing more information to be remembered (e.g. that an item was stressed) increases the chance of a confusion, or allows at least the possibility of recalling a stressed digit as unstressed, and vice versa. Our findings were that stressed digits suffered less from the confounding effects of the speech suffix, so we must assume that the acoustic analyser must be able to do more than simply compile all the relevant features into a bundle. We must allow it to make some discriminations based on phonetic evidence, and also to act on the results of these discriminations. This opens up the possibility of another modification of the original explanation of PAS.

Indeed, one may argue that maintaining information regarding the perceptual coherency and contour of the presented information is a primary function of PAS. The speech suffix, therefore, can be regarded as having a severe confounding effect on the order (as well as the identity) of the last few items. In general, PAS may be able to retain a fair stretch of the contour or some other reduced aspect of the signal, although it would not actually have specific information about the detailed microstructure of much of the waveform (the sort of information, in fact, that would enable it to characterise the items phonetically). In the experimental situation in which PAS is studied by means of the suffix effect paradigm, the particular exigencies of the experiment could force upon PAS the necessity of saving microstructural information about the items whose further processing has been artificially blocked by the arrival of the suffix.

Some tenuous confirmations of such a proposal are citable. For instance, Crowder (1972) performed a suffix experiment on lists made up from the items "ba", "ga" and "da". He found that there was little effect due to changing a speech to a non-speech suffix. An experiment on lists made up from items in which the initial consonant stayed the same but the vowel differed disclosed that PAS was more sensitive to the vowel part of the item than to the consonant part: he found the suffix effect re-instated in the latter experiment.

Now, vowels are important parts of the speech signal, not least because the information contained in the vowel sound may assist the identification of lexical items. For instance, if vowel transitions are eliminated or tampered with, confusions of temporal order occur (see Cole and Scott, 1974 for a review of evidence appertaining to this point). This is at the phonetic level. As far as prosody goes, Allen (1967) got Ss to tap to a tape loop of speech in time to the syllables, and found that the "syllable beat" of speech was identifiable with the location of the vowel onsets of the syllables (although Morton, Marcus and Frankish, 1976, report that this might turn out to be true only of words whose first consonant is a fricative). Fundamental frequency carried by the vowel is an important cue to stress (see Lehiste, 1970, chapter 6); and of course the vowels and voiced consonants of speech define the intonational contour. Finally, Noteboom (1972) reported an experiment carried out by Sliss (which was subsequently replicated by the present author). The words of a short sentence were recorded in isolation, and then cut out of the tape so that a small part of the onset and offset ofeach word was left behind: this ensured that the excised portions of the tape were fully occupied by the recorded words. When these portions were joined up in grammatical order, although the words themselves could be identified, in the words of Noteboom:

> It is as if the words were spoken by different speakers, in different corners

> of the room, and the order of the words cannot be.determined. This effect... dissapear(s)...when pauses of about 100 msec. are introduced between the words.

It was also found that rate was not a critical factor, since playing the tape at the same speed without the 100 msec. pauses as the total playing time of the tape with the pauses produced the same effects as described above. Noteboom concludes:

> The main difficulty seems to be caused by the extensive jumps in voice pitch at the beginnings and endings of words.

It should be emphasised that this interpretation does not in any way imply that the entire list is represented in PAS at any point in time. Rather, it envisages PAS as a repository for the contour of an utterance (or at least part of it). The acoustic analysis system provides feed-forward for the operations of the cognitive system on the lexical information of the utterance. It may also provide a store of what has been received not long ago in case there is a need to re-analyse any part phonetically. It is therefore only in the context of a suffix experiment that puts the emphasis on storing phonetic information relevant to the last few items in PAS. Although this capacity demonstrably exists it is not the primary purpose of PAS, according to this interpretation.

This argument is presented in the spirit of an hypothesis arising from a summary of the obtained results integrated with some of the known facts discovered elsewhere about the acoustic analysis

system. It is only tentatively presented as an explanation: further research is obviously needed in order to test the equating of PAS with a general short-term and largely prosody-orientated acoustic store for speech. It is obvious that a lot of research remains to be done on the properties and function of the PAS and acoustic analysis system, which if carried on here, would give the thesis more of the character of a monograph on short-term memory than one on speech perception would warrant. In the final experimental chapter, therefore, the investigation turns to the question of how much prosodic information is picked up by listeners from fairly long stretches of speech.

Chapter Four

In this chapter a sample of speech will be examined to discover how much information regarding syllable structure, stress, and tone group division is perceptible in the speech signal when this signal has been transformed in a way which denies a listener access to grammatical levels of processing.

Concern with grammatical aspects of speech processing has tended to minimise the rôle played by prosody in speech perception in much current psycho-linguistic theorizing. A plausible justification for this state of affairs may run something along the following lines. Since we can understand speech in which prosodic cues are reduced or absent, interest should focus on the grammatical cues in speech (see for instance, Fodor, Bever and Garret, 1975, p.344). We have been able to observe from the experiments reported in chapter two that this assumption may be an over-simplification of the problem. Chapter three reported evidence to the effect that attention may be paid to the prosodic aspects of speech before grammar is attended to. Nevertheless, it may still be objected that although prosody may indeed be helpful, in the normal course of events, a speaker does not provide these prosodic cues. This chapter, therefore, will attempt to show that for one stretch of speech at least, a speaker did provide prosodic cues in enough detail for listeners to be able to extract them independently of grammar: that is, that the perception of prosody is not attributable to some sort of "perceptual constancy" as suggested by Fodor, Bever and Garret (op. cit., pp. 299 - 301).

Bolinger and Gerstman's (1957) and Lieberman's (1967) demonstrations of the over-riding effects of pause on perceptions of stress have been taken as strong evidence to the contention that the acoustic-phonetic coding of prosody is complex and therefore not invariant (see Fodor, Bever and Garret, op. cit., p. 295). The demonstration was replicated, following the two above-mentioned sources, and the results were as follow.

The sentence in example (1) was recorded twice, once with emphasis (i.e. tonic, or in American linguistic parlance, major stress) on "light" and once on "house", producing readings corresponding to examples (2) and (3):

1. a light house keeper.
2. a light-house keeper.
3. a light house-keeper.

Example (2) was thus perceived as "a man who tends a light-house" and (3) as a "house-keeper who does not weigh much". Introducing some blank tape at the points marked with a "/" produced readings corresponding to examples (4) to (7):

4. a light/house-keeper.
5. a light-house/keeper.
6. a light/house-keeper.
7. a light-house/keeper.

Although the perception of examples (2) and (3) above corresponded fairly to intuition, paradoxically, examples (4) and (6) were perceived as (3)--"a

housekeeper who is light", and (5) and (7) as (2). That is, it made no difference as to what the position of the tonic was, the percept followed the grammatical pattern of bracketing suggested by the location of the blank tape (or "disjuncture"). Note, however, that the location of the tonic did not change mysteriously with the manipulation of the disjuncture. Example (4) sounds as if the speaker wished to draw the attention of the listener to the fact that the house-keeper is light, and example (7) that the topic of conversation is a person who tends a light-house (and not, for instance, a light-ship).

On the other hand, if the phrase was recorded on a monotone with each word produced at a regular temporal interval (see the recording procedure for experiment two), and the subsequent manipulations were done on this material, the percept once again accorded with the location of the disjuncture, but the location of the tonic was still indeterminate.

In summary, disjuncture can be regarded as a useful cue to constructs, but only in a roundabout (and unproven) way can it be regarded as a cue to stress. This roundabout way involves the listener in synthesising an intonation contour for himself from the perceived surface structure, imposing this synthesis on what his ears have already told him, and then convincing himself that he is actually hearing the overlay of his own making as part of the signal.

Lieberman (op. cit., chapter seven) also took some spectrographic measurements of an informant saying the two versions of example (1) and compared the produced disjunctures between the relevant words of these to the amount produced between "light" and "heavy" in a recording of the phrase "a light heavy-weight".

It turns out that the amount of disjuncture between "light" and "heavy" is mid-way between the amount of disjuncture between "light" and "house" in the two meanings of (1). He says:

> the relative ambiguity of the constituent structure of each phrase as derived from its words affects the degree to which the disjunctures reflect the constituent structure (p. 157).

He does not mention stress, until:

> disjuncture would manifest the constituent structure where it would otherwise not be clear from the total context of the message. The perception of the "weaker" degrees of stress always follows from the listener's application of the rules of grammar on the derived phrase marker (p. 159, my italics).

This, although no proof to the effect that changing the constituent structure of an otherwise phonologically ambiguous utterance (i.e. by adding disjuncture to a monotone reading) has been adduced. In fact, what Bolinger and Gerstman concluded their study with was:

> the disjunctures function directly to carry the information, and not indirectly as components of a hypothetical stress.

Fodor, Bever and Garret (op. cit.) must surely have misunderstood the point when they wrote:

> the introduction of a pause at the morpheme boundary between /t/ and /h/ will have the effect of converting a phonetic string with the perceived stress /light(1)#house(3)#keeper(2)/ (i.e. lighthouse keeper) into a string with the perceived stress /light(1)#house(1)#keeper(2)/ (i.e. light house keeper)...thus we have comparable silent intervals effecting contrasts in ...perceived stress (p. 296).

On the other hand, let us suppose for a moment that it did turn out that disjuncture was a cue to stress. The difficulty here would be that, providing the relationship was at all lawful, the mapping would not be simply one-to-one--i.e. a stressed item contained some acoustic event which identified it as stressed without regard to context, phonetic or grammatical--but perhaps many-to-one: for example, subject to transformational rules at the acoustic level. This would in no way diminish the theory of perceptual invariance at the level of prosodic processing (see the discussion on invariance in chapter one).

Studies from the "click" paradigm have been cited as relevant to the question of the interaction between acoustic cues and structural (grammatical) cues in sentence perception. The original "click" experiment (Ladefoged and Broadbent, 1960) contrasted Ss' ability to report the location of extraneous short bursts of noise ("clicks") in spoken sentences and strings of digits. It was found that Ss were more accurate at locating clicks superimposed on the lists of digits, thereby

indicating that we do not perceive speech word-by word. Further studies tried to isolate the controlling variables of the effect.

Fodor and Bever (1965) hypothesised that it was the location of the major syntactic boundary (MSB) that was important. They found that their Ss' ability to locate clicks accurately increased when the click was objectively located at the MSB, and also that errors tended to "migrate" towards the MSB: that is, clicks objectively located near the MSB tended to be reported erroneously as occurring at the MSB. However, it was noted by Garret, Bever and Fodor (1966) that in fact, there are also acoustic cues at the MSB which might influence the Ss' decisions. They therefore recorded sentences in which the acoustically defined boundary (or, to revert to our earlier usage, prosodic boundary, PB) was both minimised in the reading, and by cross-splicing, made to conflict with the MSB. Their experiment, therefore, was not actually concerned with evaluating the relative importance of the MSB and the PB, but rather to see whether the MSB by itself was an important factor in sentence perception. It was. A further study by Abrams and Bever (1969) used stimulus sentences constructed out of words pre-recorded in isolation. Although this study was principally concerned with reaction time to clicks, Abrams and Bever found that "71% of the click placement errors of $^{1}/_{2}$ syllable were towards the direction of the MSB" (presumably if the click was objectively placed a syllable or more's distance from the MSB it was reported erroneously, if it was reported erroneously, elsewhere).

However, when Wingfield and Klein (1971) explicitly did not maintain intonational neutrality it was found that more ear switches (they used ear switches rather than superimposed clicks) were located correctly when they objectively occurred at the PB than at the MSB, although there was no difference between these two positions when ear switch migrations were taken into account. Thus it cannot be maintained that the MSB is the only controlling variable in these studies. It is likely to be so only when the PB is deliberately minimised. Minimising the PB, however, is no control for the possibility that the PB might be an important influence on the location of clicks in the first place.

Thus so far, there is no evidence to the point that listeners actually ignore prosody when attending to prosodically normal speech. However, another study of Lieberman's (1965) has often been cited in support of the notion that:

> Lieberman (in press [sc. Lieberman, 1965--J.K.]) has shown experimentally that the perception of intonation and stress is dependent on a knowledge of the syntactic structure (Bever, Fodor and Weksel, 1965).

--a strong claim to make, especially in view of the fact that the authors allowed it to be re-published after the original study by Lieberman had been published in full. The actual evidence published by Lieberman, however, is not to this point at all. Lieberman investigated the ability of three linguists to transcribe utterances using the Trager and Smith notation system (see Trager and Smith, 1951). This

system involves assigning to each syllable a value of one to four depending on the prominence of that syllable, and juncture symbols from a set of three between some of the syllables. Lieberman's stimuli were transformed so that his Ss could not make out the words but so that the envelope amplitude, fundamental frequency, both or either, were maintained. As a control, he included the original stimuli. He concluded that:

> the phonemic pitch levels and terminal symbols of the Trager-Smith system often have no distinct physical basis"

--a remarkably less ambitious claim than the one attributed to his study by the authors cited above.

In fact, one interesting part of Lieberman's study was that one of his Ss also transcribed the pitch contour of the utterances using "a tonetic intonation". This notation marks the pitch of a syllable as either high or low, and whether it is level, falling, or rising; or a combination of the latter two. The notation also distinguishes between two sorts of junctures, one of which coincides with falling pitch levels (for a fuller description, Lieberman refers us to Stockwell, 1961). Lieberman found that this notation involved drastically fewer errors. The errors that were produced were mainly extra placements of a juncture associated with non-falling contours. Even here, this assignment of junctures was more consistent than that made with reference to the Trager and Smith system. It seems in general, that the device used by Trager and Smith involving the separation of four levels of stress from

intonational devices (which includes junctures) has been put seriously into question by Lieberman's data with regard to the acoustic basis of speech.

Of course, a trained linguist can learn a code that regulates any particular pattern of behaviour to any category of a system of grammatical classification: the issue here is whether the pattern of behaviour is under the control of acoustic or grammatical variables.

Yet it is clear from a consideration of speech that some acoustic characteristics of speech give rise to the perception of some prosodic features. Kirakowski (1973) demonstrated that at least "major stress" of a short sentence was perceptible from the sound wave rather than from a contextual or a grammatical analysis.

For this experiment, ten sentences were constructed, of the grammatical form "subject+ verb+object". For each sentence, three contexts were added. Each context drew into emphasis one particular constituent of the "target" sentence. These contexts and target sentences were recorded by naive speakers, the contexts were edited out, and the resulting productions of the target sentences were played back to listeners. The listeners' task was to indicate which of the words contained the syllable that held the "major stress". It was found that Ss performed the task with considerable accuracy.

It appears from these studies that the major outlines of the intonation of an utterance are perceptible from the sound wave alone, and do not need to be based on a consideration of the grammatical or contextual aspects of the utterance. However, it is not clear whether junctures and stress assignments of a binary nature (stressed and unstressed--se chapter one) are equally obtainable under these conditions.

The raw material for the next two experiments was a paragraph 128 words long from a novel by E. Waugh. This paragraph and some of the data collected about it in the next two experiments, is reported in the appendix. It was recorded by a female speaker with a fairly low tessitura of voice. The machine used throughout was a Revox two-channel tape recorder with a moving coil monocardioid microphone. The informant later analysed the script of the recording she had made into syllables according to her own intuitions. This was the basis for all subsequent syllable-counts.

Experiment 7

The concern was to see if listeners could segment continuous speech into smaller groups (e.g. tone groups), using acoustic criteria alone. Rees (1975) discusses certain phonetic cues that a tone group provides as to the location of its boundaries. There are four separate but potentially combinatory categories which he mentions: pause is his first. He notes that the pause marker normally coincides with the boundaries of the large syntactic units, such as the sentence or the clause, and that this is the easiest to perceive. However, it might also be that our perception of disjuncture (here equated with the acoustic feature /pause/) is not given so much by the relevant acoustic cues as by the perceptual intactness of the grammatical clauses on either side of it. There are two possibilities in this latter respect. Firstly, it may be that large stretches of speech do not exhibit /pause/ at all; and secondly that pauses of all kinds may occur at all positions within clauses, only that the within-clause pauses are ignored and the between-clause pauses are prominent (the reason why between-clause pauses would be perceived in favour of within-clause pauses in this account would be that associated grammatical cues--boundaries and suchlike--would tend to enhance the former and suppress the latter).

Method

Design. The two conditions were (1) normal direction of playback, no manipulation of signal; and (2)

reversed direction of playback to ensure that Ss were unable to understand the words of the stimulus but still had the same amount of acoustic energy. The tape was played to three Ss six times in each condition, alternating conditions. Ss task was to indicate whenever they heard a pause.

Procedure. The reversed direction tape was made by running it backwards and playing from the opposite track of the two-track machine. This procedure was checked spectroscopically to ensure that it simply reversed the speech, which it did.

Ss listened to the recording through a loud-speaker, and tapped a pencil on a desk whenever they thought a pause was occurring or had just occurred. A second tape recording was made of the experimental situation. Taps on this recording were later summarised on a printed script. The recording of the reversed condition being administered was simply reversed once again. This turned it the right way round.

Subjects. Ss were three postgraduate students in the psychology department. They were naive as to the purpose of the experiment.

Results

Even with a delay ascribable to reaction time, most taps occurred within the pause that they were intended to designate. Thus the interword

position Ss had attempted to designate were readily discernible, and a summary of the number of taps at each interword position was made, as described above. The average length of a group boundaried by taps was 7.45 syllables, with a standard deviation of 2.33 syllables. This was found to be near enough to the length of an average tone-group: Laver (1970) cites this as being "seven or eight syllables".

If disjuncture is perceived with reference to the acoustic feature of /pause/ then we should expect a high value for the correlation coefficient phi (Nunally, 1975) between the two conditions. Phi can be applied to dichotomisable variables and in our case we can dichotomise the 127 possible interword positions in which at least one tap was made and positions which received no tap at all. This dichotomy can be further cross-classified between the two conditions. Table 4.1 shows the resulting two-by-two matrix.

TABLE 4.1: Number of inter-word intervals in which one tap occurred, cross-classified between the forwards and the reverse conditions.

		forwards	
	taps	no taps	totals
backwards taps	21	0	21
no taps	2	104	106
totals	23	104	127

A value of phi is obtained from this table of .946. This indicates that the inter-word positions which received a tap in one condition of playback also received a tap in the other condition.

Discussion

This simple experiment confirms the expectation that for one sample of speech at any rate, there is an acoustic feature, /pause/, which cues the percept of a disjuncture between words. This feature is perceptible even when access to grammatical processing is denied the listener. Several observations from the data are noteworthy.

Firstly, in two cases, taps were made in the forwards version at a place for which no taps were made in the reverse version, i.e.:

8. As he stood on the verandah (7/0) calling for his boy.

(the figures in brackets should be read as "seven taps made in the forward version, none in the backwards)

9. Low at present (2/0) but with the promise of a fiery noon.

It is possible that in these two cases, and a few others in which the number of taps in the reverse version was small (i.e. four or five, of which there were two fours and two fives) that the tap in the forwards version was motivated by features other than /pause/. The effects of intonation, tempo, and grammar and semantic bond are not separable in this instance.

Secondly, all the taps were made at syntactic boundaries, and it was never the case that a minor syntactic boundary was attended to at the expense of a major syntactic boundary near at hand. However, it is not entirely clear from this data whether there is indeed a simple rule which would predict the placement of a pause from the text. Some occurrences are obvious, for instance:

10. which had taken place overnight.(18/13) The rains were over. (17/16) The boards were warm under his feet; (18/18) below the steps...

In these cases the pauses are separating sentences or large clauses. Other cases are not so obvious. For example, why is there not a pause before "of" in example (12) as there is in example (11)?

11. where before had hung a blank screen (9/8) of slatey cloud.

12. below the stone steps the dark weeds of the landlady's garden.

In general, it seems that major grammatical boundaries are marked with pauses, but that sometimes minor ones are as well.

Thirdly, it seems that response to a pause is variable in the sense that Ss did not agree with each other 100% very often. An investigation of the distribution of the scores between Ss revealed that the low non-unanimous scores were not attributable to any one subject. If the number of times a particular inter-word position was recognised as a pause is taken as an indicator of the confidence of the Ss that a pause juncture

actually occurred there, and the confidence of a response is taken as an index of the amount of information present in the stimulus which is capable of controlling the response, these variations in response may be taken as an index of the amount of disjuncture perceived between the words.

The product-moment correlation coefficient, r, was calculated for all the interword positions which have at least one tap within them in either condition between the two conditions. The resulting value, .986, is significant beyond the 1% level. It suggests that the same factors affected Ss' confidence in both conditions; or to put it another way, the degree of perceived disjuncture itself is not particularly dependent on the perceiver's hearing the words of the speech and hence the grammatical structure.

Further research is needed to examine the intonational as well as the tempo aspects of the speech wave in order to determine how far is the perception of disjuncture cued by prosodic features other than pause. A start in this direction will be made in the next experiment.

Experiment 8

The motivation for the analysis-by-synthesis model was the assumption that segmentation is impossible at any usefully small level of phonological encoding (see chapter one). Recently, new arguments have been put forward for the consideration of the syllable as an important unit in perception (for instance, see Lehiste, 1972, p. 199 et seq.). Nevertheless, it remains to be shown that fluent speech is capable of division into syllables before any operation such as lexical identification takes place.

For present purposes we will not attempt to define syllables using such terms as "boundary" and "nucleus" but in terms of an operational definition, whereby a listener simply reports how many syllables he can hear in a given stretch of utterance. We have already observed the segmentation of an utterance into units comparable with the tone-group. It remains to be asked if there are acoustic cues to a unit of organization between the syllable and the tone group: the metrical foot (see chapter one). As has been seen from Lieberman (1965) the status of acoustic cues to stress is at present not entirely clear.

As in the previous experiment, the speech will be examined for the relevant cues in experimental conditions (two in this experiment) where the S cannot make out the words of the stimulus, and a control condition where no

tampering has taken place with the signal at all. The experimental hypothesis is that syllablization and foot structure are perceptible from cues in a signal even when access to lexical and grammatical levels of decoding is rendered impossible for the listener.

Method

Design. The tape recording used for the previous experiment was divided into sixteen "segments" at points of maximum agreement between Ss as to where disjunctures had occurred. These segments were subjected to two experimental transformations: one was reversing (as before) and the other was spectral filtering which passed only the fundamental frequency contour. The control condition comprised of the segments without any manipulation performed on them. Six Ss listened to the segments, one segment at a time, on a tape loop. The control condition was presented last, to avoid the possibility of Ss associating the words of any of the segments with a fragment of acoustic information.

Ss task was to listen to the tape loop as many times as they wished and to write down a '|' sign for a stressed syllable, 'U' for an unstressed syllable, and '/' for a disjuncture between any of the syllables, such as might be given, for instance, by a silent ictus.

Procedure. For the bandpassing, a Barr & Stroud variable filter type EF 2 was used, set to pass below 250 Hz. with an attenuation slope of 72 dB per octave. This was the highest pass which left listeners unable to report any of the words or syllables of the original recording, and was established beforehand with two naive Ss who were not used again in the experiment.

The experiment used two tape-recorders: on one was mounted the tape containing the separated segments, and on the other was set a tape loop. Each segment in turn was transcribed via a lead between the machines. Since Ss found the task of attending to the tape loops of reverse speech very difficult, these were slowed down to half the speed of the original with a correction for pitch made by the Lexicon VARISPEECH. Agreement between Ss as to the number of syllables was, if anything, worse when the tape loop in this condition was played at the normal speed. Ss wrote their responses on sheets of paper; they were all tested at the same time.

Subjects. The eight Ss were all postgraduates or members of staff of Edinburgh University.

Results

The analysis is reported in two parts. First the data will be examined for evidence of syllablization, and then the agreement between conditions about the

sequence of stressed and unstressed syllables will be examined.

Syllablization. The average number of syllables for each segment was compared between the three conditions and with the syllable count as given by the informant. The standard deviations for the averaging across Ss within each segment ranged from zero to 3.8 with an average at about 1.4. Table 4.2 gives the inter-correlation matrix for the four pair combinations using Pearson's r.

TABLE 4.2: Intercorrelation matrix for number of syllables reported per segment.

	transcript	normal	filtered
reversed	.834	.860	.907
filtered	.957	.969	
normal	.995		

All correlations are significant beyond the 1% level. A similar matrix was also computed for each S, and although there was more variability in the obtained coefficients, the pattern was still the same and all the correlations were still statistically significant. It seems, therefore, that the experimental manipulations did not affect the Ss' ability to report the number of syllables per segment.

The perception of stress. A measure is needed that will reflect how closely Ss are in agreement about the pattern of stressed and unstressed syllables in each segment. Since not all Ss reported exactly the same number of syllables, a measure which depends on

absolute positional information (e.g. the serial method of scoring experiments 5 and 6) will tend to give too many instances of disagreement where in fact little disagreement obtains apart from errors of omission or comission. The following procedure is an attempt to devise a metric that is sensitive to relative, not absolute, position.

Suppose the reports of two Ss for the same segment are considered a syllable at a time. If both Ss are in agreement about whether the nth. syllable is stressed or not, then we may pass on and consider the (n + 1)th. syllable. If there is a disagreement between the two Ss, one of two things has happened: either one of the Ss has missed out reporting a syllable (or the other has reported one too many); or one of the Ss has reported a syllable as stressed when the other reported it as unstressed. Choosing between these two possibilities will have important consequences for how the rest of the two reports will be scored, but there is unfortunately no external criterion which will help us to decide--that is, we cannot appeal to a record of the "correct" answer.

What we can do is to consider both possibilities in turn, and analyse the rest of both reports twice, once as if the omission/comission hypothesis were correct, and once as if the disagreement hypothesis were correct. These two analyses will eventually produce different figures for the total number of disagreements about the subsequent syllables.

A conservative criterion for which hypothesis was correct then becomes the one which led to fewest subsequent disagreements.

For instance, suppose that two reports turned out as follows (for this imaginary data, 1 represents a stressed syllable, and 0 an unstressed syllable):

Subject A: 100101

Subject B: 10101.

The two reports agree until the third syllable. If we hypothesise that S(B) has misidentified this syllable, we have to assume that either he has missed out the fourth syllable as well, which would bring the total disagreement score up to two, or that he has misidentified all the remaining syllables and missed out one at the end, which brings the total disagreement score to four. On the other hand, if we hypothesise that S(B) has missed out the third syllable, and gone on to report the fourth syllable instead, then the subsequent disagreement score is zero, and the total disagreement score is only one, which accords with our intuitions when we examine the two sets of reports. Going back to the method, comparing the total disagreement scores for the two possibilities at syllable three, we see that the most parsimonious assumption to make is the omission/comission one, which accords with the intuitive interpretation. The method cannot tell us which S has made the error, nor can it tell us whether the error was comission or omission.

The example discussed above takes the case of only one disagreement between Ss: when there are more than two or three points of disagreement the calculations become complex and are best left to a computer program.

If every possible pair of reports for the segment is treated in this way, the resulting total disagreement score will reflect the total amount of disagreement between all the Ss. If there are six Ss, there will be (5+4+3+2+1=) 15 comparisons to be made in each segment in each condition. Table 4.3 shows the total disagreement scores for all 16 segments in all three conditions, with means and standard deviations. These totals appear to be normally distributed from an examination of the number of cases within one, two, and three standard deviations from the mean. What is important here is not whether these totals are significantly different from each other, but whether they are significantly higher than the chance expectation.

Two methods were used to derive values for the chance expectation. One was a "Monte Carlo" method in which the experiment was replicated by a computer program that simulated six Ss responding with a random sequence of stressed and unstressed syllables to the proportions of occurrence of these syllables in the actual data in segments whose mean syllable length and standard deviation was defined by the average over the experiment.

TABLE 4.3: Totals of disagreement scores.

	Conditions:		
Sequence no.	Reversed	Filtered	Control
1	115	120	63
2	79	61	61
3	28	22	5
4	51	42	25
5	52	45	32
6	60	42	55
7	54	73	59
8	47	24	30
9	63	69	67
10	61	63	35
11	26	32	15
12	49	38	34
13	59	58	49
14	49	45	30
15	50	60	32
16	103	22	81
Means	59.125	55.375	42.062
Standard deviations	22.08	24.56	20.07

The other was an "a priori" method whose reasoning was as follows. If we consider six fictitious reports of one syllable, the summed disagreement score for the fifteen possible pairings of that result range from zero (no disagreements) to nine (three syllables reported as stressed and three as unstressed). The average disagreement score is five. Thus the total expected disagreement score will be five times the average number of syllables reported for the segment.

The two methods agree pretty closely for the estimated chance frequencies. The Monte Carlo method tends to give larger values for the amount of disagreement expected by chance; however, the difference between the two methods is not statistically significant (t=1.5373, d.f.=30). Since the fluctuations observed in the Monte Carlo method's results could be due to sampling error, expected frequencies will therefore be estimated by the "a priori" method.

For each condition, the total expected score (by chance) was calculated for each segment and it was found that only the control condition's scores were significantly above chance levels. The results of these calculations are summarised in table 4.4.

TABLE 4.4: Differences between observed total disagreements and expected (chance) total disagreements for the three experimental conditions.

	Observed means	Expected means	Sandler's A	
reversed	59.125	49.05	.1098	NS *
filtered	55.375	53.70	2.6720	NS
control	42.062	57.90	.0838	.01

d.f. = 15

* although this value of A is significant, it shows that the observed mean is greater than the expected (chance) mean. We cannot therefore reject the null hypothesis in this case.

Discussion

We can conclude with some certanity that there is information in the acoustic signal specific to the number of syllables that make up a particular segment (e.g. a tone-group) of speech. What seems to be in doubt is Ss' abilities to characterise each syllable as stressed or unstressed. One possibility is, of course, that Ss were operating a three-level stress system and therefore asking them to make two-level judgements brought about a lack of agreement as to where the middle level of stress should fall. This explanation does not, however, square with the results of the control condition, which showed that when Ss were scoring the forwards untransformed tapes they were more unanimous than a chance response hypothesis would have predicted.

The Ss seem to agree as to the total number of syllables there should be in each segment, but do they also agree as to the number of stressed syllables? When the correlations between the average number of stressed syllables reported per segment are computed between the three conditions (see table 4.5) the resulting intercorrelation matrix shows that the agreement between conditions is substantial.

TABLE 4.5: Intercorrelation matrix for stresses.

	control	filtered
reversed	.854	.838
filtered	.950	

All three correlations are highly significant. Since the number of syllables (stressed and unstressed) also correlates highly across conditions (see table 4.2) this implies that the real difficulty in the experimental conditions is not so much the identifying of a particular syllable as stressed or unstressed, but of representing the order of stressed and unstressed syllables. In the experimental conditions, the Ss are unsure of the order, and this is manifested by the high disagreement scores within these conditions; in the control condition, Ss are relatively more unanimous about the order. However, Ss are always in agreement about how many stressed syllables there should be. This interpretation is supported by observations from the Ss:- they found that it was easier to attend to the stressed syllables first, and then to work in the unstressed syllables in around them.

The other interesting aspect of the data not yet brought out concerns the reporting of junctures within this experiment. Some of the segments contained interword positions about which there was only a modicum of agreement from experiment seven between Ss as to whether a pause had occurred or not. Although the reporting of disjunctures in the present experiment was not emphasised, there were some segments in which all the Ss did actually mark a disjuncture within three or four syllables of each other. Bearing in mind the range of the standard deviations for the number of syllables reported per segment, this is a fair amount of tolerance.

But first, in the two cases where a disjuncture had been reported in the forwards version of experiment seven, although nothing had been reported in the reverse version, the pattern of responses in experiment eight remained the same; i.e. we must accept that these disjunctures were not recognisable with reference to the sorts of information presented in the experimental conditions of either experiment. These were:

13. As he stood on the verandah (E7:7/0,E8:4/2/1) calling for his boy.

(The data for experiment seven are reported as before, first the number of taps in the forwards condition, then the number of taps in the reverse condition. The data for experiment eight are reported in similar fashion: first the number of "/" signs within four syllables in the untransformed condition, then in the filtered condition, and lastly, in the reverse condition).

and:

14. low at present (E7:2/0,E8:2/0/0) but with the promise of a fiery noon.

Now, if we accept that a disjuncture reported in the filtered condition is indicative of a disjuncture that is best cued intonationally, whereas disjunctures reported in the reverse condition are cued mainly by the presence of pauses, in three cases, a strong intonational cue was noted:

15. villas and farms and hamlets (E7:6/4,E8:3/6/4) gardens and crops.

16. rolling green pastures (E7:10/4,E8:3/6/0) dun and rosy terraces.

17. He slowly became aware of the transformation (E7:6/2, E8:5/6/3) which had taken place.

This data is reported in a spirit suggestive of further research and the lines along which it could take place, rather than as a definitive statement.

General discussion

One criticism that may be made about both experiments is that the sorts of manipulations the speech signal underwent actually attenuated the acoustic cues to the perception of finer degrees of disjuncture and stress. Although this must remain a possibility, several arguments mitigate against it.

Firstly, with respect to disjuncture, only a small proportion of the cases actually involved a disagreement between the forwards and the reverse condition (two out of 127). If the reversing had systematically eliminated some sorts of acoustic cues to disjuncture we would have expected this discrepancy to have been a lot larger since it must be admitted that such cues would be present in the control (unmanipulated) condition.

On the other hand, the three occasions when Ss were unanimous about the presence of a disjuncture in the filtered condition of experiment eight, were not points at which there was no indication at all of a disjuncture in the reversed condition of experiment seven: that is, although disjunctures may be cued by pauses there are also perhaps more effective cues in the intonational contour which may even co-incide with pauses. In fact it seems to be the case that rather than giving too little acoustic information about disjuncture, we have been able to observe our informant giving more than enough on

occasion. It remains to be seen exactly how much prosodic redundancy of this sort is typically given by speakers.

Secondly, with respect to stress, it may be argued that filtering and reversing eliminated the crucial cue to stress perception: for instance the rather sharp attenutation slope could have been systematically lopping the tops off the intonation peaks at each stressed syllable and that a reverse temporal order renders such intonational or other (for example, second formant) cues to stress unaccessible. There are two answers to this objection. Firstly, if Ss were responding with stressed syllables randomly, one would expect a much lower--a statistically insignificant--correlation between the number of stresses perceived in each condition. Indeed, if this were the case, Ss could even have perceived rising-chopped-falling contours on one (objective) syllable as two syllables and a juncture. The fact that in the filtered condition the average number of syllables reported per segment correlates highly with the transcript is an important argument against such a possibility, as is the high correlation observed between the average number of stressed syllables reported per segment between the two experimental and the control condition.

Secondly, the objection does not do justice to the wealth of cues that a speaker may use to produce a stress. Lehiste (1970, Chapter four)

summarises a number of experiments which attempted to discover the relevant cues to the perception of stress using short stimuli such as bi-syllabic words or phrases. The important fact which comes out of her summary is that at least three, if not four sorts of cues are important: fundamental frequency, amplitude, duration, and voice quality (see also chapter one). Fundamental frequency, admittedly, is seen as the most readily perceptible, although all three have been shown to be effective. To suggest that either experimental transformation actually eliminated all three (or four) sorts of cues is not too plausible.

What has emerged clearly from both experiments, however, is the fact that there are a number of prosodic cues which can be picked up directly from the sound wave without recourse to grammatical or semantic levels of representation. On the other hand, there is not so much support for the notion that unstressed syllables are processed in the same manner as stressed syllables.

These considerations will be taken up in greater detail in the last chapter, which attempts to integrate the research findings presented in this and the preceding two chapters with a theoretical model of speech perception.

Chapter Five

In this concluding chapter, a model of speech perception will be presented, and the results obtained in the three previous experimental chapters will be integrated within the model. The model itself attempts to describe the functional sub-parts of the process of speech perception, and to relate these parts to each other.

Firstly, however, an introductory section will define the scope of the model; introduce the sort of explanation that is going to be proposed with a critical examination of some of the pertinent literature; and present a framework for discussion.

Introduction

The perception of speech can be regarded as two operations: those of derivation and representation. By representation will be taken to mean the problem of how is information arriving at the listener's ears represented in a form that will enable him to perform such diverse tasks as repeating it to someone else, checking the discrepancy between it and a picture, judging the truth or falsity of it, or relating it to some other information the listener has acquired, not necessarily through the ears. In other words, how can we characterise "what a person knows of the meaning of a sentence" (Clark, 1976, p.11), once he has comprehended it or derived it?

By derivation will be taken to mean the operations that are entailed before representation can be considered: that is, how is the gap between

the sounds entering at the ears and the representation of those sounds as described in the previous paragraph met. The question to which the following sections will address themselves is: what should the derivation procedure look like?

These two problems are separable in that we can certainly talk about what the representation will look like without having to consider problems of derivation (see Clark, op. cit.). It follows, though not with certainty, that we can ask what the derivation will look like without having a specific theory of representation in mind. Since there is no accepted body of knowledge which is recognised by every psychologist concerning either of these problems just now, it would therefore be as legitimate to consider representation without derivation as it is to consider derivation without representation.

In fact, any theory of either derivation or representation will have to satisfy the criteria of objective verification, and it may be argued that once this has been met, the problem of translating from one theory to another is trivial. However, it is doubtful whether the state of our knowledge about speech perception at this time is sufficient to enable the formulation of any theory specific enough to make the task of devising a translation worth while. It is hoped that the theory as it will be presented will be general enough so that a lot of previous findings about speech perception

can be integrated within it yet capable of being narrowed down to a more precise statement (at some later date) by the testing of specific predictions arising from it. In the third section of this chapter, the results of the experiments reported previously in this thesis will be accommodated within the theory.

It was noted in chapter one that perhaps the most fruitful hypothesis concerning the way speech is processed is to assume that the "abstract performative grammar" (c.f. Watt, 1970) is composed of a loosely-ordered series of rules. The important questions then become, firstly, what sorts of rules are there (i.e. what are the component sub-parts) and secondly, how are these rules integrated in the act of perception (i.e. what are the inter-relationships between parts). A modest amount has been written attempting to answer these two questions about such rules of "perceptual mapping".

In 1967, Neisser discussed some "cues to phrase structure" (pp. 259 - 267) where he examined the cue-value of prosody, function words and affixes towards recovering the surface structural representation of a speech input. With regard to function words and affixes, his discussion amplifies the later discussion by Clark and Clark (1977) although his remarks about prosody have not been taken up in the context of a model for speech perception.

Bever (1970) discusses twelve rules which describe strategies of getting from what he calls "actual sequences" (spoken or written evidence) to "internal structures" (meaning). He divides them into: segmentation, functional labelling, and semantic strategies.

Firstly, with regard to segmentation strategies, he says:

> failure to separate the correct basic segmentation into sequences that do correspond to underlying structure sentences could seriously degrade comprehension (p. 288, op. cit.).

He claims to show support for:

> the existence of a perceptual strategy of isolating lexical sequences that correspond directly to underlying structure representations (p. 289, ibid).

In itself, this is ambiguous, but the context indicates that Bever intends "separate out by referring to underlying structure" rather than "such lexical sequences that correspond to underlying structure representations": that is, his "perceptual segmentation" strategies depend on the very analysis they are designed to help elicit being determined beforehand. We shall refer to this apparent paradox later.

As for functional labelling strategies, they assign

> the internal structural relations which bind the constituent phrases in each internal sentence (p. 295, ibid).

Perceivers use, according to Bever, strategies based

on (a) probabilistic structural features; (b) semantic information; and (c) "knowledge of potential structure underlying specific lexical items" (p. 295, ibid). Semantic strategies "combine lexical items in the most plausible way". Bever notes

> the most likely semantic organization among a group of phrases can guide the interpretation of sentences, independently of and in parallel with perceptual processing of the syntactic structure (p. 297, ibid).

That is, "functional labelling strategies" and "semantic strategies" can be considered as alternatives rather than hierarchical processes. However, functional labelling strategies based on semantic information and knowledge of potential structure underlying semantic information are surely themselves also describable as strategies which "combine lexical items in the most plausible way": in other words, it is difficult to see either from Bever's definitions, or from his examples (see especially pp. 299 - 303) exactly what he intends as the critical difference between his semantic and functional labelling strategies.

Clark and Clark restrict their discussion to what they call "syntactic" and "semantic" rules, preferring to avoid the discussion of perceptual segmentation altogether. Their syntactic and semantic rules correspond to Bever's "functional labelling" and "semantic" strategies, although the correspondence is not always very clear: for instance, on page 57 (op. cit.) they seem to be

saying that syntactic strategies enable a listener to "build and connect propositions in an interpretation for the whole sentence"; with semantic strategies, however, "listeners are assumed to work from the interpretation a sentence must be conveying"(my italics) . Later, instances they give of semantic strategies do not compel the position that listeners are working from the overall interpretation to search for cues in the signal; for example:

> Strategy 8: using content words alone, build propositions that make sense and parse the sentence into constituents accordingly (p. 73, ibid)

goes surely from a consideration of the possible semantic relationships each word may enter into, to inferences about the scope of surface structure constituents.

However, supposing that the syntactic and the semantic approaches are separable according to the criteria cited, a strategy such as

> Strategy 10: look for definite noun phrases that refer to entities you know and replace the interpretation of each noun phrase by a reference to that entity directly (p. 76, ibid)

presupposes that a recognition of noun phrases has already taken place. Although they say that listeners "most probably use some mixture of the two approaches" (p. 57, ibid) it seems from examples such as these that listeners must first use one and then the other.

Kimball (1973) discusses five principles of labelling and bracketing sentences (all five of which are also discussed by Clark and Clark) and relates

these principles within a "top-down" model, given as the first principle:

> parsing in natural languages proceeds according to a top-down algorithm (p. 20, op. cit.)

that is, analysis works from the topmost "S" node to the terminal nodes of the input. Processing takes place after labelling and bracketing have been accomplished, as given in principle seven:

> When a phrase is closed, it is pushed down into a syntactic (possibly semantic) processing stage and cleared from short term memory (p. 38, ibid).

He does not discuss what these processing rules will look like. His perceptual rules are limited to segmentation and labelling operations.

Kimball's reasons for accepting a "top-down" model and rejecting a "bottom-up" or even some compromise are never made very clear (although on p. 21 it appears that some compromise is necessary). Principle seven rejects the possibility that in analysing sentences semantic features may play any rôle at all in deciding what the segmenting and labelling should be. Since his principles two to five presuppose that each word has been identified fully as to meaning and grammatical function--c.f.:

> Principle Three (New Nodes): The construction of a new node is signalled by the occurrence of a grammatical function word (p. 29);

--it is not clear why he should so arbitrarily restrict himself in this regard.

In summary, the approaches discussed do not distinguish between operations of segmentation, labelling, and computing meaning to a sufficient extent, although the authors cited obviously intend that these processes are proper sub-parts. The rules all assume that each lexical item has already been sufficiently well recognised for grammatical and semantic processing to take place.

It is clear, therefore, that (a) an attempt should be made to define the above-mentioned operations clearly enough that they may be kept separate, and that (b) the way these operations interact should be laid down in more formal terms than heretofore. Also, (c) that the stages of processing necessary before lexical identification can take place should be made part of the model. In fact, it is important to realise that any segment of speech may be considered not only as a symbol (or sequence of symbols) conveying certain sorts of semantic and grammatical information, but also as a sound with a certain acoustic structure, potentially integrated within a larger sequence of sounds, all of which are capable of carrying important information to the perceiver (see the discussion in chapter one).

On another tack, two seemingly different approaches to explaining perception in general have long been contrasted. One approach embodies the "information processing model" which can be summarised with reference to a tube. Speech enters down one end of the tube, and meaning

comes out of the other end. The workings of the tube are complex, which implies that information does not necessarily flow through it in one direction all at the same pace--of which more below.

The alternative approach is the "constructional" theory, which can be likened to a ring. Information arriving at the senses is placed on the ring, and passed round and round, becoming more "refined" on each passing cycle. The latter day equivalent of the tube model is surely Morton's logogen model. Ulric Neisser, once at least partly a tube man (q.v. Neisser, 1967, pp. 196 - 198) is now decidedly a ring man (Neisser, 1976, p. 18).

The difference between these two modes of explanation is sometimes fancifully distinguished as "passive" (tube) vs. "active" (ring); but the difference is perhaps more apparent than real. Neisser (1976) overstates his case when he allots a central rôle--the cardinal rôle--in his theory to the device of anticipation. Anticipation has to be defined broadly enough so that (a) most events the system will have to deal with can be anticipated, for un-anticipated events cannot be processed by this model (apart from the paradox of anticipating an un-anticipated event); and (b) the anticipation is hardly ever wrong, in order to correspond with observations of introspection. Neisser acknowledges both objections (ibid). Both these statements about

anticipation can, perhaps be stated more credibly in terms of tube models. Thus (a) would be stated: events for which the tube has no analysers cannot be analysed; and (b): the operation of a certain set of analysers will involve the operation of certain other sets, and the disuse of certain yet other sets.

As mentioned above, the tube model need not be stated as baldly as to imply a straight-through process from one end of the tube to the other. On the contrary, the effects of feed-back and feed-forward in speech perception have become so well established that it is almost banal to re-iterate them. As discussed in relation to Morton's model, and as we shall see again, later, these principles are easily compatible with the tube model.

In fact, what appears to be a fundamental difference between sorts of possible explanations that can be offered on closer examination turns out to be a difference simply of vocabulary. In these concluding paragraphs the point has been made that there is no real difference between a model proposed under the auspice of either, provided the model is constrained and relates to the empirical evidence. The subsequent sections of this chapter will use the tube model as a frame-work. Interested partisans may make the translation into ring terms. If translation fails, there may be opportunity for dialogue.

The model

Data-base and input

The perceiver has to combine information from two sources in order to understand speech: his data-base, which can be identified with the "abstract performative grammar" (c.f. Watt, op. cit.), and the incoming evidence. The data-base may be considered as a set of rules which can be applied in a certain order to the evidence. A simple distinction may be made as to the incoming evidence: there is the information from the evidence currently being analysed, and the information from the state of the rest of the system as a result of operations on incoming evidence other than that which is currently being analysed.

Incoming evidence

It will be taken as axiomatic, and corresponding with our intuitions as listeners, that incoming evidence consists of segments which can be ordered hierarchically and linearly . Segments in order hierarchically, will, in the situation of perfect communication, be analysed in parallel; segments in order linearily will, in these circumstances, be analysed in sequence. In this case, there will always be agreement in the conditions of feed-forward and feed-back between and within parallel and serial processes: however, speech perception does not characteristically take place in ideal acoustic conditions, and therefore

the conditions of feed-forward and feed-back (among others) assume paramount importance in the model as interactions. This topic will be discussed in detail below.

Thus at any moment, the listener is in receipt of information from a number of segments of different hierarchical order. He also has information from previously analysed segments, context, and expectation.

The primary evidence for speech perception, however, which will be the theme of the following discussion, is the information arriving at the ears. This may also be supplemented with visual information when such is available--e.g. gestures, movements of the lips, etc., and rather more fancifully, from evidence reaching the listener from the other senses. This is what may be called the "synaesthetic component" of speech perception. It is important to mention it, for a complete theory of speech perception will have to account for the contribution of the synaesthetic component, although for the present discussion, there is no room to expand on it.

In the case of the perfect communicational situation, information from all these sources of input evidence will be congruent. Interesting questions begin to arise when this is not the case. Which source of evidence will then be given precedence over the others depends therefore, on (a) the specificity of each of the sources, and (b) the

the weight of the evidence attached to the other sources. Morton's logogen model is seen as an embodiment of these proposals with regard to lexical identification.

Three plausible operations of derivation

Each operation is seen as divided into two "phases". In general terms, the first phase performs the work of recognition, and the second, that of integration of the hierarchical segments.

Segmentation. The units of speech will be considered for the moment as: phonemes, syllables, words, constituents, sentences, and possibly also paragraphs. Whether or not there are actually more or less than these is for present purposes immaterial, so long as the units of perception can be seen to stand in a hierarchical relationship to each other. The choice of these particular units was made because for each of them, at some time or another, it has been claimed that they have a perceptual reality which has made them distinct from the adjoining segments of the same level. From this it follows that there is a perceptual segmentation operation performed on these very units. The second phase of the segmentation operation attempts to integrate the segments into the larger segments of which they form a part; that is, to establish a rudimentary bracketing. In order to do this, the segmenter must be able to do three things: (1) detect the start of the segment; and (2) detect the end of the segment.

Slightly less obviously, it must also be able to (3) register when it is in the middle of processing a segment, and to differentiate this from the situation when it has stopped processing a segment.

Identification with respect to function. In this operation, such aspects of segments like syllables, words and constituents which do not depend on an exhaustive analysis of their meaning are attended to. In many cases, it will be seen, this is tantamount to an identification of the grammatical function the particular segment plays within the larger segment of which it is a part. In the second phase of this operation--integration--the output from segmentation is confirmed (or not, as the case may be) by a preliminary structural analysis (for instance, a tentative labelling of the bracketing) of the segments arising from the first operation (segmentation).

Identification with respect to meaning. On the basis of the first two operations (and context, expectation, prior analyses, etc., see incoming evidence) certain relationships between segments have been suggested. The primary meaning of each word or clause is here accessed and in the second phase investigated as to how it fits in with other segments of the same hierarchical level, and higher-order segments. It is here, perhaps, that the problems of derivation cease and the problems of representation begin to take over. Sentences or paragraphs are assimilated into the representation the listener is making of the state of things.

Interactions

Interactions can be regarded as decisions taken at one stage of representation or analysis which necessitate additional computation at another stage. Feed-forward and feed-back are therefore two sorts of interactions between operations that can take place. These two suggest a third, namely that of abeyance; by which is meant the suspension of a decision until other information appears which might be relevant to the making of the decision. The resolution of abeyance, therefore, necessitates some sort of feed-forward or feed-back, but these two latter conditions do not of themselves presuppose abeyance: the so-called "garden path" sentences, for instance (e.g. "the horse raced past the barn fell") are instances of feed-back on decisions previously made.

Strength of the model

A weak claim is that, for any one segment, these three operations should be kept logically distinct although in a real-time model, they may take place simultaneously. A stronger claim is that, barring interactions, these operations must be applied in ascending order on any segment of a given size that is capable of being treated by all three operations.

The strong form is preferable to the weak, in that the weak is a counsel of despair and an acknowledgement that the scientific study of the

process of perception is impossible within the present framework. If the weak form is adopted, the perception device becomes a "black box" in which anything can happen. If, on the other hand, it turns out that evidence or sheer common sense obliges us to connect everything with everything else indiscriminately of the conditions that prevail at the time of perception by means of interactions, the strong form will have disappeared, and we will eventually be left with the weak. But it will not have been for want of trying. In the two following sections, then, firstly the arguments concerning the logical separateness of the operations will be considered, and then those concerning possible process distinctions, as far as the data gathered in the earlier chapters will allow us.

Evidence

Logical distinctions

Wingfield's paradox, alluded to earlier, is an argument about the logical distinction between segmentation and further processing. To postulate that segmentation is based on meaning is unsatisfactory if we also argue that meaning is based on segments that have been correctly grouped (c.f. Wingfield, 1975, p. 146). The resolution of the paradox entails our postulating that identification and segmentation are two distinct processes. However, it is perfectly possible to contemplate procedures that segment and identify at the same time, so we have to look for other arguments to support the real-time separation of these two phases of perception.

A second argument is based on the infinite generative capacity of grammars. In order to be adequate, a grammar must have such a capacity, although an infinite sentence must remain a purely theoretical entity, perhaps in the realm of science-fiction. From this, it follows that for some units of perception at least, the segmenter which relies solely on identification can never be certain at any given part of the segment whether this part is the concluding element of the segment or not. As Kimball (op. cit.) points out, what is necessary is a look-ahead feature in such a segmentation-and-identification routine. This may be adequate in a device such as a compiler of

a computer language (see Knuth, 1965) but in speech, the existence of such a look-ahead feature is impossible, unless we are prepared to admit that speech perception involves ESP. What it means in practice is that for some kinds of segments, at least, the routine will not know that a segment has ended until the first element of the following segment has been considered and analysed.

According to this argument, one would expect that from studies which examine the amount of processing done at various points in the perception of units such as clauses, processing contingent on the end of a clause will be deferred to the start of the next. This does not seem to be the case. Using the length of reaction time to an extraneous event or a particular sound in speech as an indicator of the amount of time the perceiver has to spare for activities other than the processing of speech, such studies have characteristically shown that reaction time increases towards the end of the phrase, and decreases shortly thereafter, to a significantly lower value at the start of the adjoining clause.

As with the first argument, this does not actually tell us anything about the order in which the operations may be carried out: it simply tells us that we have to consider them as two separate activities, which may either occur one after the other, together, or together with interactions.

There are two arguments that may be applied to the logical distinction between functional and meaning identification. The first depends on the contention that a single word in isolation may have a multiplicity of meanings, and that an isolated clause may have a variety of interpretations depending on its context. From this, it follows that some initial constraints on the possible range of meanings of a segment before it enters meaning identification is desirable. Of course, sources of information other than those present in the speech signal may make their influence most manifest at this stage, but our ability to comprehend without undue difficulty sentences presented fairly devoid of context indicates that something from within the segment is constraining our choice of meaning. Conversely, words which sometimes function as auxiliaries and sometimes as main verbs are on the whole correctly interpreted by listeners (and readers) with little effort and few "false starts". Once again, something within the sentence is constraining the choice of meanings: identification of grammatical function could indeed be doing just that.

Secondly, our ability to understand "Clockwork Orange" type sentences (e.g. "the gloopy malchicks scatted razdazily to the mesto") on the one hand, and to fill in the function words and affixes to strings of content words like "girl give flowers soldier" argues that the operations of functional identification and meaning identification

can at least in principle be considered as separate operations.

Order of operations in real time

The bulk of the empirical work presented in this thesis can be considered as support for the following contention, which can be stated in two parts. Firstly, segmentation takes place before functional identification in a real-time sense, although there is an interaction between these two processes. Secondly, the main data for segmentation (or the speech evidence for it) is contained in the prosodic aspects of speech. Arising from the experimental data, two other points can be discerned. They are, firstly, that some of the information relevant to functional identification can be gleaned from the prosodic aspects of speech, and in particular, from information gained in the segmentation phase. Secondly, the functional locus of the segmentation operation, and possibly also of the functional identification operation as well, is pre-categorical (in terms of Morton's model): that is, before the activities of accessing the meanings of words and phrases are considered. These three contentions will be treated separately in the three sections which follow.

Segmentation by prosodic cues. The way functional identification routines work, according to Bever (op. cit.) and Clark and Clark (op. cit.) is by examining the nature of the class of each word, and

by building up plausible grammatical structures on this basis. Thus if the words of the stimulus are kept the same, the same functional identification routines will be called and the perceiver will make the same segmentations, regardless of other possible cues to segmentation in the signal. However, using ambiguous sentences which have at least two possible meanings, experiment one showed that in many cases, manipulating the prosodic pattern changed the segmentation. This result argues for the dependence of functional identification on segmentation routines based on prosodic analysis.

In fact, there are at least two sorts of ambiguity that prosody can affect. One sort, called by Kirakowski (1975) "floating" ambiguities, consist of three grammatical elements, x, y and z, not necessarily placed in that order in the surface structure, where y is ambiguous in that it can form a construct with either x or z, but not both. Whichever element it can form a construct with does not however alter the grammatical function of the ambiguous element. That is, an adverbial phrase, for instance, can go with either the verb phrase x or the verb phrase z; in neither case does it change its function as an adverbial phrase. An example of such a sentence is: "the paper he wrote quickly limited analyses of the problem", where"quickly" is the ambiguous item, "wrote" and "limited" are the two verbs.

More interesting are ambiguous sentences in which the ambiguous element (y) undergoes a

syntactic change of identity when transferred from the x constituent to the z. Thus y may be adverbial when it is segmented with x, but adjectival when segmented with z. This sort of ambiguity was called "stable changing" by Kirakowski (ibid). An instance of "stable changing" ambiguity in experiment one is the following example: "because of her injury that summer she stayed at home". "That summer" is the ambiguous element, which can either agree with "her injury" or "stayed". Since prosody successfully cued the disambiguation in this case, it follows that the identification routine was dependent on the output from the segmentation routine before it could make any identification at all.

However, not all the stimuli from experiment one were so successfully disambiguated. When a clear effect did not show itself for any one example, it was not clear whether this was due to the lack of a sufficiently powerful prosodic cue, or whether it was due to the cueing of a requisitely powerful identification strategy in the teeth of a prosodic cue. If the latter were the case, there were too many instances of this happening in the experiment for any case to be made for its general applicability. However, the data from this experiment on its own cannot tell us which of the two alternatives was responsible for the effect.

Experiment two provided data that enables us to resolve this problem. If the identification strategies of Bever (op. cit.) are powerful enough

on their own, they would not need any prosodic cues. Comparison between the conditions with normal and reduced prosody of the second experiment disclosed that this was not the case: perception was "seriously degraded" (to borrow Bever's phrase--see above) in the absence of appropriate cues to segmentation given by prosody. However, prosodic cues to segmentation were only found to be relevant if functional identification could take place on the data: comparison of normal and reduced prosody conditions in the ungrammatical passage showed no effect of prosody. Another way of stating this might be to say that adding or taking away some of the prosodic information from speech which is not susceptible to the processing of meaning does not seriously degrade or enhance its intelligibility. This interesting finding will crop up again in the context of the discussion of experiment eight, where it will be entered into in greater detail.

To counter the objection that cues to phrase structure are apparent only in artificially-constructed material (although the naive informant condition of the first experiment would suggest that this was not generally the case, anyway, although more seriously it could be argued that since there were only about two points in each sentence where the phrase marker could go, Ss were attending to minimal cues), experiment seven (and eight) examined the existence of cues in the acoustic signal to phrase boundaries that were independent of grammatical information, could occur anywhere, and were produced in the course of a long stretch of speech. The high

reliability of the obtained measures suggests that such cues are readily available to listeners under these conditions.

Still, the evidence from the first two experiments and experiments seven and eight could be interpreted as evidence for an interaction between segmentation and functional identification in line with the "weaker" theory. In order to attempt to demonstrate that segmentation can occur before identification, experiment six was devised. The results, interpreted in terms of Morton's model, suggest that the prosodic features of tone-group final contour is important before the speech evidence ever gets to a stage where the lexical items can be identified, hence certainly before meaning identification, and arguably before the identification of function, although this will be discussed below. It is important to note that use of Morton's model in no way implies an uncritical acceptance: as will be seen later, integrating it into the general model of speech perception necessitates some fairly radical alterations of interpretation. On the other hand, in discussing the importance of results in suffix-effect experiments, use of the model is unavoidable, simply because it is the most complete explanation given to date of not only the original effect but also the many related suffix effects.

To conclude this section, a simple demonstration of the effective nature of prosodic patterns for suggesting segmentations can be made along the lines

suggested by Braine (1964). The laryngograph processor (Fourcin, 1975) can be used to record only the larynx frequency of laryngographically processed speech on tape. This recording is fairly even in intensity, but is a faithful reproduction of the fundamental frequency ("intonation contour") and the rhythmic aspect of the speech signal. If white noise is superimposed on such a signal, listeners can be persuaded that they are listening to speech which has been band-passed and mixed with white noise, and indeed, if one is aware of the deception, the impression is hard to shake off. Under these conditions, with some coaxing, Ss can be persuaded to repeat back what the speaker is "saying": if the sample is put on a tape-loop, each repetition of the loop brings with it greater conviction of the rightness of the guess. The constituents and number of syllables Ss produce, and sometimes also the order of function and content words, although not the same in meaning, are certainly very often strikingly similar to those of the original recorded message. Since this effect has only been tried casually on a few Ss in the laboratory, it must retain the status of anecdote until a fuller treatment of it is possible.

Functional identification. The problem of separating function from meaning identification is to devise two plausible sets of procedures, which do not work on the same aspects of the speech wave. Furthermore, functional identification should not rely on exhaustive information about the rôle each segment can play in the larger segment of which it is a part

since this would defeat the object of the separation. Functional identification should be based on readily-detectable features of speech; the data-base for meaning identification is the full range of semantic features associated with each segment and its potential combinations, constrained by outputs from functional identification.

A solution to this problem harks back to Neisser's discussion on the information in function words and affixes, mentioned earlier (Neisser, 1967). As Neisser pointed out, and as is apparent in the "Clockwork Orange" example, information from these parts of speech enables the listener or reader to compute a labelled bracketing of the input, without any recourse to the semantic features of the content words (which in the Clockwork Orange example, are very poorly represented unless one is conversant with Burgess' neo-Russian slang).

A simple demonstration of the priority of functional identification is as follows. Suppose we take example (1) as a fairly normal-sounding English sentence:

1. The manager duped many customers with his system.

Some approximation to the meaning of this sentence can be deduced from a consideration of the content words alone, as in example (2):

2. manager dupe many customers system.

On the other hand, we could fairly easily "fill out" example (3) to produce a bracketing and labelling something like the one given by example (4):

3. The xxxed xxxs xxxed xxxly with the xxx

4. The duped customers argued harshly with the manager.

Note that in example (4) the order of words, considered simply as a sequence of content or function words, is the same as in example (1). If the 'xxx' symbols in example (3) are replaced by the content words, in their actual order, of example (2), the result is a string in which the function word and affix cues are in conflict with the meaning suggested by a consideration of the content words alone, as in example (5):

5. The managered dupes manyed customersly with the system.

Most readers encountering example (5) attempt to analyse the string by taking the information from the function words and affixes to produce a phrase structure bracketing and labelling similar to that of example (4). Had function and meaning been heavily interactive to the extent of being best treated as one single process, readers would as often take a bracketing similar to that of (1) as to (4) when reading (5); had they been working from a semantic strategy in the first place, they would take a bracketing similar to that of (1) in preference. Neither of these latter alternatives actually seems to be the case, so we are left with the hypothesis that functional identification actually comes first, constraining the choices open to the meaning identification routines. Obviously, this is another experiment that would have to be repeated in more controlled conditions than the rather informal ones in force in the present situation; but the effect

is a robust one, and presented with example (5) on a piece of card all of seven Ss, on asked to describe the bracketing and labelling they initially perceived on seeing the example, opted for a version similar to that of example (4). None remotely considered example (1) as a possibility.

The argument that this section will try to establish from the data previously presented runs along the following lines: content words normally receive a stress in English, and usually, function words are left unstressed (except in special cases of emphasis which might themselves be revealing). Thus a simple heuristic to a listener could be: try to build up a labelling of the bracketing from segmentation, using unstressed syllables alone, around the stressed syllables. On the basis of this, engage the appropriate logogens for the content words.

Experiment three manipulated the perceived stress on the words making up the stimuli, and it was found that when the stress was reduced, making content words and function words appear on a monotone, and then at equal temporal intervals with regard to the syllables, comprehension was impaired. Once again, the comprehension of ungrammatical sequences similarily treated was not terribly much impaired from one condition to another, arguing that the effect is specific to information that will undergo further treatment by the listener (function and meaning identification).

Experiment eight, by contrast, investigated the amount of information conveyed in normal speech with regard to the identification of stressed and unstressed syllables, with and without information about the meaning of the speech. It was found that listeners were not in agreement about the actual sequence of stressed and unstressed syllables when they could not hear the meaning of the stimulus, although they did agree as to the absolute number of both present in each stimulus. This finding has since been replicated by Daw (1977) in a M.A. thesis, using a slightly different method.

Daw went on to test whether this finding was due to the fact that Ss could not actually discriminate between stressed and unstressed syllables, or whether it was due to Ss' inability to represent a discriminated sequence in the absence of further information (i.e. grammatical). She concluded that the problem was with Ss' representations rather than with their discriminations. This is suggested both by the finding about the correlations between conditions on the number of stressed and unstressed syllables reported, and also by a re-consideration of experiments two and three.

That is, the perceptual system, although it may be functionally analysable into separate components, and although it may be possible to draw a flow-path of information through these components, acts as a unit rather than as a bundle of separate entities. If a blockage occurs at

any point in the stages of analysis (e.g. inability to identify syllables, or constituents) then much of the previous computations (correspondingly: identification of stress, segmentation of clauses) becomes unavailable to conscious processes. This "holistic" aspect of the speech perception device may help to explain some other previously puzzling observations about perception: for instance, the old argument about the "speech mode" of auditory perception being qualitatively different from the "non-speech mode". According to the holistic hypothesis, this could have been expected: although some linguistic analysis has taken place on sounds which approximate to speech but are obviously not speech, when the identification process fails at some stage, all the benefits from previous processing are lost as well, and the listener is left with the non-speech like sound.

Perception of speech being an active process, however, such a holistic system may sometimes be deceived by suggestions that a given stimulus is or is not speech, as in the laryngographic demonstrations mentioned earlier. It has a parallel in that a S can be sucessfully persuaded that distorted speech (for instance from an old and worn tape or from a severely band-passed recording) is not speech but a collection of squeaks and buzzes.

Morton's speech/non-speech switch now becomes a reality in such a model, not as a

bundle of neurones which either pass or refuse admittance to stimulation reaching the ears (not that Morton ever intended such an anatomical application--see Morton, 1970--but it could have been perfectly feasible as an anatomical feature), but as a functional property of the perceptual system, echoed within the structure of the system at many levels. The study of what is retained and what is not under such conditions of "holistic rejection" may disclose further knowledge of the dynamics of the system.

Finally, in this discussion of the importance of the distinctions between stressed and unstressed syllables, experiment five disclosed that in fact stressed syllables are treated by the early stages of the processing system in a qualitatively different manner from unstressed syllables. The precise mechanism which has to be added to PAS is not yet quite clear, yet on grounds of parsimony it would be advantageous to make this mechanism responsible for the Salter effect (Salter et al., 1976) as well: not only stressed syllables, but also syllables which are in some ways phonetically different from the preceding context are given priority of treatment by this mechanism within PAS. Given that function words and affixes are a relatively small subset of all legal syllable combinations in English, and that their frequency of occurrence is much greater that that for syllables from outside this category, the speculation is as follows.

This mechanism responsible for the Salter effect is responsible for segmentation and functional identification. Functional identification may be accomplished after a differential response to high-frequency stressed syllables and low-frequency unstressed syllables. It can operate without prosodic cues to segmentation, but at reduced efficiency (see experiment four). Thus both segmentation and functional identification are seen as pre-categorical activities on information held in PAS; meaning identification is seen as a post-logogen activity. Morton himself posits a feed-forward line from PAS to the cognitive system; duration of stay in PAS may also enable the cognitive system to "have another look" at the contents of PAS, should this be required.

Meaning and prosody. The greatest contribution prosody may make to the computation of meaning is surely as a result of its crucial rôle in determining the segmentation and labelling of speech input, as discussed previously. The studies reported in earlier chapters and summarised in the preceding sections have a lot to say about this aspect of the contribution of prosody to perception. They have relatively little to say about the contribution of prosody to the computation of meaning directly, yet prosody does seem to play an important rôle here as well.

One crucial aspect of intonation might be to help establish the "given-new" distinction in the representation process (see especially Clark and Clark, ibid, chapter three), and it is interesting to see how Halliday (1967) considers intonation relates to this problem.

Each tone-group which contains one tonic element conveys one major information point (in the case of a tone-group with two tonic elements the situation is more complicated and will not be dealt with here). The selection of which item is to be made tonic can be regarded as the distribution of information units (see Halliday, op. cit., pp. 21 and 22).

When the tonic does not occur at the last lexical item of the tone group, it is classified as a "marked" tonic, according to the system. Marked tonics specify either of both of two things. One is which element or elements of the tone group are to be considered contrastive; and the other is what is to be considered as given.

Consideration of which element is the tonic in a tone-group, can, therefore, assist in a "semantic" strategy such as Clark and Clark's strategy 14 (Clark and Clark, op. cit.):

> Look for given information to precede new information, unless the sentence is marked otherwise (p. 79).

Some evidence for the rôle of intonation in this respect comes from experiment three, in the comparison of the "normal" and the monotone "foot-timed" conditions: between these two for the grammatical sentences, a decrement in correct reporting was observed. It is reasonable to suppose that the stage of meaning identification was awaiting and depending on intonation to produce some sort of information relevant to the given-new distinction in these cases. It is not clear from the circumstances whether this was simply because of the isolated, out-of-context, nature of the experimental stimuli, or perhaps more interestingly, indicative of a fundamental processing strategy in all situations of listening to speech. Evidence from Kirakowski (1973) discussed previously (see chapter four) might suggest that speakers do provide information of this sort in their speech, and that listeners are equipped to pick it up, thus indicating the second of the two alternatives.

However, this seems to be as far as the evidence from the present data can take us in this direction: the reader is referred to Crystal (1975; see especially chapter one) for a fuller discussion of the sorts of information prosody and intonation in particular may have on listeners' representations of sentences.

Concluding remarks

The model presented in the second section of this chapter has been discussed and related to the empirical evidence from the experimental part of the thesis in the present (third) section. If the model has any general validity, it should suggest testable hypotheses; and some of these have been outlined in brief. It should also be compatible with other psycho-linguistic and linguistic findings, and this is a task at which the scope of the thesis means it can only hint.

With respect to the workings of the model, it seems that segmentation and the identification of function are two operations carried out while the speech evidence is relatively un-categorized (i.e. with respect to lexical and grammatical identification); and the presented evidence has been interpreted to the effect that these two operations could be ordered in real time within a mechanism operating on the contents of pre-categorical storage.

The arguments presented in this chapter and at the end of chapter three claim that prosodic information is most important at these stages of processing, and that consequently, the true nature of PAS is that of a device for maintaining such prosodic information.

More generally, prosody has been seen to affect every part of the tri-partite model. This,

more than anything else, is the theme of the entire thesis: prosodic information is crucial to the perception of speech, and any model which does not recognise this fact is open to self-contradiction or the accusation of solipsism (or as in the case of the analysis-by-synthesis model reviewed in chapter one, both). This does not mean that the inclusion of prosody automatically exonerates a model from these faults, but it is our contention that it gets us nearer the truth.

Appendix I
Experimental Materials

Experiment 1

1. She told them where some red coats were.

redcoats: soldiers.

red coats: coats which are red.

2. Because they run around so much, in the evening, rats are very sleepy.

rats run in the evening.

rats are sleepy in the evening.

3. The paper he presented quickly limited analyses of the problem.

he presented the paper quickly.

the paper limited the analyses quickly.

4. The vandals broke the window of the little green house.

a house, painted green.

a greenhouse (such as is used for plants).

5. This book I found accidentally shed light on the topic.

the book shed light on the topic for me by accident.

I found the book accidentally.

6. Pavlov fed her dog biscuits.

dog-biscuits (biscuits for dogs).

gave her dog some biscuits.

7. They took the plane to another city.

they boarded the plane which was going to another city.

they flew the plane to another city.

8. I asked her how old George was.

how old he was.

how "Old George" felt.

9. He told me to write it in strictest confidence.

to write it in strictest confidence.

he told me in strictest confidence.

10. His friend laughed derisively at the church.

when he was at the church, he laughed.

he derided the church.

11. Beethoven, knowing how great symphonies sound, composed nine.

Beethoven knew how good a symphony sounds.

Beethoven knew what all great symphonies should sound like.

12. Hannibal sent troops over a week ago.

he sent them, more than a week ago.

he sent them over, a week ago.

13. They ran out of the boxes almost immediately.

they left the boxes.

they used up all the boxes.

14. He told me to go without any hesitation.

he told me to go immediately.

he told me without hesitating.

15. They soon learned how good meat tastes.

that meat tastes good.

the taste of good meat.

16. He built the piano in that corner.

he built it in that corner.

the piano stood in that corner.

17. Because of her injury that summer she stayed at home.

she stayed at home that summer.

she got injured that summer.

18. He was telling her baby stories.

stories to her baby.

stories for babies (baby-stories).

19. He saw that beautiful Indian dancing.

beautiful dancing done by Indians.

a beautiful Indian, who was dancing.

Experiment 2

Introductory (control) paragraph:

The description of persons who have the fewest ideas of all others are mere authors and readers. It is better to be able neither to read nor write than to be able to do nothing else. A lounger who is ordinarily seen with a book in his hand is, we may be almost sure, equally without the power or inclination to attend either what passes around him or in his own mind. Such a one may be said to carry his understanding about with him in his pocket, or to leave it at home on his library shelves.

Experimental paragraph:

There is indeed a degree of stupidity which prevents children from learning the usual lessons, or ever arriving at these puny academic honours. But what passes for stupidity is much oftener a want of interest, of a sufficient motive to fix the attention and force a reluctant application to the dry and unmeaning pursuits of school-learning. The best capacities are as much above this drudgery as the dullest are beneath it. Our men of greatest genius have not been most distinguished for their acquirements at school or at the university.

Experiments 3 and 4

Grammatical sentences:
Youngsters like the spring.
He was running home quickly.
A large meal was ordered.
No-one came to see her.
Who could blame the painter?
He neatly swept the floor.
She was combing her tresses.
Can you drive this lorry?
I did tidy my room.
Are these people your friends?
He came at supper time.
Trusting him is hardly safe.
Who will come at Christmas?
The blossoming flowers were nice.
She nearly broke her cup.
Give aid to the needy.
He gave her a sickly smile.
Why are you so polite?
Morning is the best time.
The hunter came back home.

Ungrammatical strings:
swept he the neatly floor
she was tresses her combing
who the could blame painter
to no-one came her see
room did my tidy I
she nearly her cup broke
spring youngsters the like

needy aid the to give
was ordered meal a large
home the hunter back came
supper at time he came
drive can lorry this you
home quickly running he was
friends your people are these
is safe him trusting hardly
smile her gave sickly a he
why so polite you are
at christmas will who come
time the morning is best
were flowers the nice blossoming

Experiments 7 and 8

As he stood on the verandah, calling for his boy, he slowly became aware of the transformation which had taken place overnight. The rains were over. The boards were warm under his feet; below the steps, the dark weeds of the landlady's garden had suddenly burst into crimson flower; a tropic sun blazed in the sky, low at present, but with the promise of a fiery noon, while beyond the tin roofs of the city, where before had hung a blank screen of slatey cloud, was now disclosed a vast landscape, mile upon mile of sunlit highland, rolling green pastures, dun and rosy terraces, villas and farms and hamlets, gardens and crops and tiny stockaded shrines; crest upon crest receeding to the blue peaks of the remote horizon.

Appendix II
Additional Calculations

TABLE A1: Analysis of variance on the data from experiment five, serially scored.

Source	SS	d.f.	MS	F	
Nonadditivity					
nonadditivity	411.8	1	411.8	1.29	NS
balance	267560.1	839	318.9		
Analysis of variance					
A (subjects)	151732.5	35	4335.2	13.59	.01
B (conditions)	15212.2	3	5070.7	15.89	.01
C (serial positions)	547643.8	8	68455.5	214.58	.01
A x B	64286.0	105	612.2	1.92	.01
A x C	205479.4	280	733.9	2.30	.01
B x C	47031.3	24	1959.6	6.14	.01
A x B x C	267971.9	840	319.0		

Since we have been unable to reject the hypothesis of nonadditivity in the data, the three-way interaction term can be used instead of the error term.

TABLE A 2: Analysis of variance on the data from experiment five, scored by the one-either-way method.

Source	SS	d.f.	MS	F	
Nonadditivity					
nonadditivity	182.2	1	182.2	0.55	NS
balance	280221.9	839	333.9		
Analysis of Variance					
A (subjects)	129612.4	35	3703.2	11.09	.01
B (conditions)	8337.4	3	2779.1	8.33	.01
C (positions)	333173.0	8	41646.6	124.76	.01
A x B	45589.4	105	434.2	1.30	.05
A x C	200553.7	280	716.3	2.15	.01
B x C	63874.0	24	2661.4	7.97	.01
A x B x C	280404.2	840	333.8		

Since we have been unable to reject the hypothesis of nonadditivity in the data, the three-way interaction term can be used instead of the error term.

Glasgow's data (1952) re-analysed

This is a re-analysis of the data presented by Glasgow for the prose conditions only with regard to the contribution of his two sub-groups to the total score.

TABLE A3: Mean comprehension scores.

	Group I	Group II	Both	s.d.	N
Good intonation	31.6	32.5	32.1	9.72	113
Monopitch	29.9	27.4	28.4	8.2	113
Score decrease	1.7	5.1	4.7		

NOTE: Data taken from Glasgow's tables I and II, op. cit., p. 67. Since Glasgow does not tell us otherwise, it is assumed that he split his total sample of 226 Ss equally between the two groups.

The standard error of the difference between means can be worked out on the assumption that the two groups were perfectly matched (since Glasgow himself mentions this fact). The standard error of the difference between means is computed as 1.167. We can now test, using the t distribution, whether the observed differences between Good intonation and Monopitch styles is significant for <u>both</u> groups.

TABLE A4: Value of t comparing Good intonation and Monopitch.

Group	t	significance
I	1.456	NS
II	4.370	.005
Both	4.027	.005

All the computations followed the procedures suggested by Runyon and Haber (1968). The results clearly show that group II is responsible for the significance of the difference of the combined scores, although group I's scores are in the predicted direction. A more appropriate statistical treatment would have been a repeated-measures analysis of variance, but there is not enough data reported for this.

Bibliography

Abercrombie, D., 1967, Elements of General Phonetics, University of Edinburgh Press, Edinburgh.

Abrams, K., and Bever, T.G., 1969, Syntactic structure modifies attention during speech perception and recognition. Q.J.Exp.Psychol., 21, 280-290.

Allen, G., 1967, Two behavioural experiments on the location of the syllable beat in Conversational American English. Studies in Language and Language behaviour, Univ. of Michigan progress reports, 4 2-179.

Bever, T.G., 1970, The Cognitive basis for linguistic structures. In: Hayes (Ed), Cognition and the Development of Language, Chap.9. John Wiley, New York.

Bever, T.G., Fodor, J.A. and Weksel, W. 1965, Theoretical Notes on the Acquisition of Syntax: a critique of 'contextual generalization'. Psychol. Rev., 72, 467-482.

Bever, T.G., Garret, M.F. and Hurtig, R., 1976, Projection Mechanisms in Reading, or when the journal review process fails. J.Psycholing.Res., 5, 215-226.

Bolinger, D., (Ed), 1972a, Intonation. Penguin, Middx.

Bolinger, D., 1972b, Accent is predictable (if you're a mind-reader). Lg., 48, 633-644.

Bolinger, D. and Gerstman, L.J., 1956, Disjunctures as a cue to constructs. Word, 13, 246-255.

Bond, Z., 1971, Units in Speech Perception. Working papers in Linguistics, 9, viii-112. Computer and Information Science Research Centre Technical Report Series, The Ohio State University, Columbus, Ohio.

Braine, M., 1964, On learning the grammatical order of words. Psychol.Rev., 70, 323-348.

Chistovich, C.A., Aliakrinski, V.V. and Abulian, V.A., 1960, Time delays in speech repetition. Voprosy Psikologia, 1, 114-120.

Chomsky, N., 1965, Aspects of the theory of syntax. The MIT press, Cambridge, Mass.

Chomsky, N. and Halle, M., 1968, The Sound Patterns of English, Harper and Row, New York.

Chomsky, N. and Miller, G., 1963, Introduction to the formal analysis of natural languages. In: Luce, R.D., Bush, R.R. and Galanter, E. (Eds), Handbook of Mathematical Psychology, vol.II, 269-321. John Wiley, New York.

Clark, H., 1970, Semantics and Comprehension. Mouton, The Hague.

Clark, H. and Clark, E., 1977, Psychology and Language. Harcourt Brace Jovanovich, Inc., New York.

Crowder, R.G., 1969, Phonic interference and the prefix effect. J.V.L.V.B., 8, 302-321.

Crowder, R.G., 1967, Prefix effects in immediate memory. Canad.J.Psychol., 21, 450-461.

Crowder, R., 1972, Visual and Auditory Memory. In: Kavanagh, J. and Mattingly, I. (Eds), Language by eye and by ear, 251-275, MIT Press, Cambridge, Mass.

Crowder, R.G. and Morton, J., 1969, Precategorical acoustic storage (PAS). Perc. and Psychophys., 5, 365-373.

Crystal, D., 1975, The English Tone of Voice. Edward Arnold, Bristol.

Cutler, A., 1976, Phoneme-monitoring reaction time as a function of preceding intonation contour. Perc. and Psychophys., 20, 55-60.

Dallett, K.M., 1964, Effects of Redundant Prefix on Immediate Recall. J.Exp. Psychol., 79, 368-370.

Darwin, C., On the dynamic use of prosody in speech perception. In: Cohen, A. and Noteboom, S. (Eds), Structure and Process in Speech Perception. Springer-Verlag, Berlin.

Daw, H., 1977, The Perception of Linguistic Stress. M.A. thesis, Edinburgh University, Psychology Department.

Derwing, B., 1973, Transformational Grammar as a theory of language acquisition. Cambridge University Press, Cambridge.

Diehl, C.F., White, R.C. and Satz, P.H., 1961, Pitch Change and Comprehension. Sp. Mon., 28, 65-68.

Figueroa, J., 1978, On the limitations of iconic memory as structural system. J.Exp. Psychol. (general), in press.

Fodor, J.A. and Bever, T.G., 1965, The psychological reality of linguistic segments. J.V.L.V.B., 4, 414-420.

Fodor, J.A., Bever, T.G. and Garret, M.F., 1974, The Psychology of Language. McGraw-Hill, New York.

Fourcin, A.J., 1975, Laryngograph analysis of vocal fold vibration. In: Ventilatory and Phonatory control mechanisms : an international symposium, ed. B.Wyke, 315-333. Oxford University Press.

Garcia, E., 1976, Some remarks on "Ambiguity" and "Perceptual Processes". J.Psycholing.Res., 5, 195-213.

Garret, M., Bever, T.G., and Fodor, J., 1966, The active use of grammar in speech perception. Perc. and Psychophys., 1, 30-32.

Gibson, J., 1966, The Senses considered as Perceptual Systems. Allen and Unwin Ltd., London.

Glasgow, G.M., 1952, A semantic index of vocal pitch. Sp.Mon., 19, 64-68

Halle,M. and Stevens,K.,I962, Speech recognition : A model and a program for research. IRE transactions on information theory,IT-8,I55-I59.

Halliday,M.A.K.,I967, Intonation and Grammar in British English. Mouton, The Hague.

Kaplan,R.M., I972, Augmented transition networks as psychological models of sentence comprehension. Art.Intell.,3,77-I00.

Kavanagh,J.F. and Mattingly,I., I972,Language by Ear and by Eye. MIT Press, Cambridge, Mass.

Kimball,J.,I973, Seven principles of surface structure parsing in natural language. Cognition,2,I5-47.

Kirakowski,J.,I973, Some effects of semantics on suprasegmental structure. M.A.thesis, Edinburgh University, Psychology Department.

Kirakowski,J.,I975,"The classification and disambiguation of graphically ambiguous sentences." Report to the Speech Communication Seminar, Edinburgh, June, I975.

Kirakowski,J., (in preparation), The sizes of lists and suffix effects.

Kirakowski,J., and Myers,T.,I975,"The effect of intonation on message intelligibility". Paper presented at the Spring meeting of the British Acoustical Society, Nottingham, April, I975.

Kirakowski,J., Vance,P., and Macnamee,M., (in preparation) Active mechanisms in PAS.

Kirk, R.E., I968, Experimental Design: procedures for the behavioural sciences. Brooks/Cole, Belmont, California.

Knuth,D.E.,I965, On the translation of languages from left to right. Info. and Control,8,I-35.

Lackner,J.R., and Tuller,B., I976, The influence of syntactic segmentation on perceived stress. Cognition,4, 303-307.

Ladefoged,P.,'and Broadbent,D.E.,I960, Perception of sequence in auditory events. Q.J.Exp.Psychol.,I2, I62-I70.

Land,E. and McCann,I970, Lightness and Retinex theory. J.Opt.Soc.Am.,6I,I-II.

Laver,J.,I970, The production of speech. In: J. Lyons (Ed), New Horizons in Linguistics, Pelican, Middx.

Lehiste,I., I970, Suprasegmentals. MIT Press, Cambridge, Mass.

Lehiste,I.,I973, Phonetic disambiguation of syntactic ambiguity. Glossa,7,I07-I22.

Lehiste,I.,I974, The Units of Speech Perception. In: J.Gilbert (Ed), Speech and Cortical Functioning, Academic Press, New York.

Leonard,L.B.,1973, The rôle of intonation in the recall of various linguistic stimuli. Lang.and sp., I6,327-335.

Liberman,A.,1970, The grammars of speech and language. Cog.Psychol.,I,30I-323.

Lieberman,P.,I964, Some effects of semantic and grammatical context on the production and perception of speech. Lang.and sp.,6,I72-I87.

Lieberman,P.,1965, On the acoustic basis of the perception of intonation by linguists. Word,2I, 40-54.

Lieberman,P.,1967, Intonation,Perception and Language. MIT Research Monograph No.38, MIT Press, Cambridge, Mass.

Longuet-Higgins,C.,1976,Perception of Melodies. Nature,263, 646-653.

Mackay,D.G.,1966, To end ambiguous sentences. Perc.and Psychophys.,I,426-436.

MacKay, D.G. and Bever, T.G., 1967, In search of ambiguity. Perc.and Psychophys., 2, 193-200.

Martin, J.G., 1972, Rhythmic (hierarchical) versus serial structure in speech and other behaviour. Psychol.Rev., 79, 487-509.

Massaro, D., 1970, Retroactive interference in short-term recognition for pitch. J.Exp.Psychol., 83, 32-39.

Massaro, D., 1972, Preperceptual images, processing time and perceptual units in auditory perception. Psychol.Rev., 79, 124-145.

Mehler, J. and Carey, P., 1967, Rôle of surface and base structure in the perception of sentences. J.V. L.V.B., 6, 335-338.

Miller, G., 1956, The magical number seven, plus or minus two. Psychol.Rev., 63, 81-96.

Miller, G., 1962, Decision units in the perception of speech. IRE transactions on information theory, IT-8, 81-83.

Mills, C.B. and Martin, J.G., 1974, Articulatory organization in the prefix effect. Perc. and Psychophys., 16, 309-314.

Morton, J., 1964, A preliminary functional model for language behaviour. International Audiology, 3, 216-225.

Morton, J., 1970, A functional model for memory. In: D.A.Norman, (Ed), Models of Human Memory, Academic Press, New York.

Morton, J., and Chambers, S.M., 1976, Some evidence for "speech" as an acoustic feature. Br.J.Psychol., 67, 31-45.

Morton, J., Crowder, R.G. and Prussin, H.A., 1971, Experiments with the stimulus suffix effect. J.Exp.Psychol.Monograph, 91, 169-190.

Morton,J. and Holloway,C.M.,I970, Absence of cross-modal "suffix effects" in short-term memory. Q.J.Exp.Psychol.,22,167-176.

Morton,J., Marcus,S. and Frankish,C., I976, Perceptual Centers (P-centers). Psychol. Rev.,83,405-408.

Myers,T.F., I973, Speech Processing Asymmetry, PhD thesis, University of Cardiff.

Nash,R.I970, John Likes Mary More than Bill. Phonetica,22,I70-I88.

Neisser,U.,I967, Cognitive psychology. Appleton-Century-Crofts, New York.

Neisser,U., I976, Cognition and Reality. W.H. Freeman and Co., San Francisco.

Noteboom,S., I972, A brief survey of some investigations into the temporal organization of speech. IPO Annual Progress Report, 7,I7-29, Eindhoven, Holland.

Ohman,S.,Numerical Model of Co-articulation. J.A.S.A., 4I, 3I0-320.

Rees,M.,I975, The domain of isochrony. Edinburgh University Department of Linguistics, Work in Progress,8,I4-28.

Runyon,R., and Haber,A., I968, Fundamentals of Behavioumal Statistics. Addison-Wesley, Reading, Mass.

Salter,D.,I973,, Shadowing at one and two ears. Q.J.Exp.Psychol.,25, 549-556.

Salter,D.,I975, Maintaining recency despite a stimulus suffix. Q.J.Exp.Psychol.,27,433-443.

Salter,D., Springer,G., and Bolton,L., I976, Semantic coding versus the stimulus suffix. Brit.J.Psychol.,67,339-35I.

Scholes,R.J.,I97I, On the spoken disambiguation of superficially ambiguous sentences. Lang. and Sp. I4,I-II.

Siegel, S., I965, Nonparametric Statistics. McGraw-Hill Book Co. Inc., New York.

Stockwell, R.P., I96I, A review of Kindon's The groundwork of English intonation (I958), Int.J. Amer.Ling., 27, 278ff. Cited in P.Lieberman (I965).

Stowe, A.N. and Hampton, D.B., I96I, Speech Synthesis with prerecorded syllables and words. J.A.S.A. 33, 8I0-8II.

Thomas, S.H., I969, Effects of monotonous delivery on intelligibility. Sp.Mon., 36, II0-II3.

Toner, H., I975, An investigation of the effect of intonation in resolving ambiguity. M.A. thesis, Department of Psychology, University of St.Andrews.

Trager, G.L., and Smith, H.L., I95I, Outline of English Structure. New York.

Wanner, E., I968, Do we understand sentences from the outside-in or from the inside-out? Daedalus, I02, I63-I84.

Warren, R.M., I970, Auditory Illusions and Confusions. Sci.Am., 233, 30-37.

Watt, W.C., I970, On two hypotheses concerning psycholinguistics. In: Hayes, J.R. (Ed), Cognition and the Development of Language, I37-220, John Wiley and sons, inc., New York.

Winer, B.J., I97I, Statistical principles in Experimental Design, (2nd. Ed.) McGraw-Hill Kogakusha, Ltd., Tokyo.

Wingfield, A., I975, The intonation-syntax interaction: prosodic features in perceptual processing of sentences. In: Cohen, A. and Noteboom, S. (Eds), Structure and Process in Speech Perception. Springer-Verlag, Berlin.

Wingfield,A;, and Klein,J.F., I97I, Syntactic structure and acoustic pattern in speech perception. Perc. and Psychophys.,9, 23-25.

Winograd,T., I972, Understanding Natural Language. Academic Press, New York.